Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

# Open Knowledge and Learner Model: Mathematical Representation and Applications as Learning Support Foundation in EFL

**Kensuke TAKII[a*], Changhao LIANG[b] & Hiroaki OGATA[b]**
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*kensuke.takii96@gmail.com

**Abstract:** In elementary education, knowing the learner's knowledge state is an essential but difficult task in today's educational settings. Learning Analytics (LA) research has successfully estimated learners' states by visualizing and analyzing their activities. Still, it was difficult for traditional LA to explore learners' knowledge states due to the lack of information about learner-knowledge relationships. Open Knowledge and Learner Model (OKLM), a universal learner model that can connect learning activity logs and knowledge models, has been developed as an LA tool that serves as a foundation for various learning support systems. Although a conceptual proposal of the OKLM framework and its applications have been provided, it has not yet been mathematically formulated. In this paper, we formulate OKLM mathematically and introduce the concept of "relative proficiency" as a knowledge state. Targeting English as a Foreign Language (EFL) learning, some of the applications of OKLM to specific learning supports, including knowledge analogy, modeling learning materials, and material recommendation, are provided. In the future, we will verify the effectiveness of OKLM using actual data and improve the model based on further validation. Moreover, OKLM will be able to broaden its context other than EFL learning. This research will contribute to opening new possibilities for LA.

**Keywords:** Learning analytics, Open knowledge and learner model, relative proficiency, learning support foundation

## 1. Introduction

Understanding a learner's knowledge state is crucial for effective educational experiences, especially in elementary education. This task is challenging in today's educational settings, where the ratio of teachers to learners is often imbalanced. With the rise of Learning Management Systems (LMS), Learning Analytics (LA) (Ferguson, 2012) has gained popularity as a method to estimate learners' states by analyzing their activities on e-learning platforms. However, traditional LA struggles to determine which knowledge-related actions a learner has taken based solely on learning activity logs.

To address these issues, the Open Knowledge and Learner Model (OKLM) was developed. OKLM, a universal learner model by Takii et al. (2023), automatically connects a knowledge model, linking knowledge items from learning materials with learning activity logs. While OKLM has been conceptually proposed and applied in some contexts, it still lacks mathematical formulation, which is essential for its broader application. This paper addresses this gap by formulating OKLM mathematically and introducing "relative proficiency" as a key concept within the model. We also present examples of how OKLM can be applied to design English as a Foreign Language (EFL) learning aids, potentially advancing the development of sophisticated learning tools.

## 2. Previous Work

### *2.1 Representation of Learner Knowledge*

Many studies are attempting to represent the learner's knowledge state mathematically. The knowledge space theory (Falmagne & Doignon, 2011) models the learner's state of knowledge based on the relationship between a set of problems and their solutions by representing the structuring of knowledge and the progression of learning. Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1995) uses Bayesian estimation to calculate the probability that a learner will answer each question correctly and infer the degree of knowledge acquisition. Item Response Theory (IRT) (Hambleton & Swaminathan, 2013) measures subject characteristics and the difficulty and discrimination of the assessment items based on their responses to a set of assessment items. However, all the above theories and methods are based on the learner's response to a particular item in question and do not estimate the knowledge state based on the learner's usual learning behavior. This study provides a theory for estimating knowledge states from learners' responses to specific items and their usual learning behavior. To achieve this, LA, which handles educational big data, accumulating learners' learning behavior, is used as a means.

### *2.2 Learner Models in Supporting EFL Learning*

A Learner Model (LM), one of the methods for representing learner knowledge, is a computer model of the learner's state and is indispensable for configuring learning and educational support systems. In the Open Learner Model (OLM) (Bull, 2020), the system makes the learner's own LM available to the learner. It allows the learner to modify it as needed, thereby increasing the accuracy and learning effectiveness of the model. The integration of LA and OLM research has long been desired (Bodily et al., 2018), and this study can be seen as a practical implementation of that desire.

LMs are also used in EFL learning. Zou and Xie (2018) developed an explicit LM for personalized vocabulary recommendation, modeling learners' vocabulary size and proficiency. Hsu (2008) also constructed vocabulary recommendations to increase learners’ motivation by modeling their interests and recommending suitable reading lessons. Takii (2022) developed an LM representing English skills for an English material recommender system. EFL learning support using LM extends beyond vocabulary to grammar learning. Aiming to provide undergraduate EFL students with individual learning paths of grammar, a Moodle plugin developed by Zabolotskikh et al. (2021), although not mentioned as an LM, automatically detects learners’ English skills and learning styles automatically. Huang and Zhu (2021) proposed an English material recommendation using a knowledge graph, which managing learners’ attributes and knowledge regarding various aspects of English, including vocabulary and grammar.

While these LMs supporting EFL learning target specific applications like recommendations and learning path detection, OKLM can be applied to other areas of EFL learning support. Several application scenarios of OKLM will be shown in section 5.

## 3. OKLM Framework and Its Mathematical Formulation

### *3.1 Open Knowledge and Learner Model (OKLM)*

Figure 1 shows the framework of OKLM. OKLM is “a learner model that can manage and track which knowledge items are covered by each learning activity by linking the learner’s daily learning logs to a knowledge map generated according to the learning materials” (Takii et al., 2023). OKLM consists of two parts: a knowledge model, which represents the knowledge items contained in the electronic learning resources, and a learner model, which is generated from the usage logs of the LA system and various learning resource browsing systems. The knowledge model is a graph structure that represents the knowledge items to be learned, and the graph structure expresses the syntactic or semantic relationships among the knowledge items. Each node in the knowledge model stores information on learning activity log data collected from the LA system and other learning resource browsing systems. OKLM does not depend on the learning resources but only on the knowledge model in the learning materials. Therefore, the learning resources from which the knowledge model is constructed can be in any format, such as PDF, PowerPoint, or video.

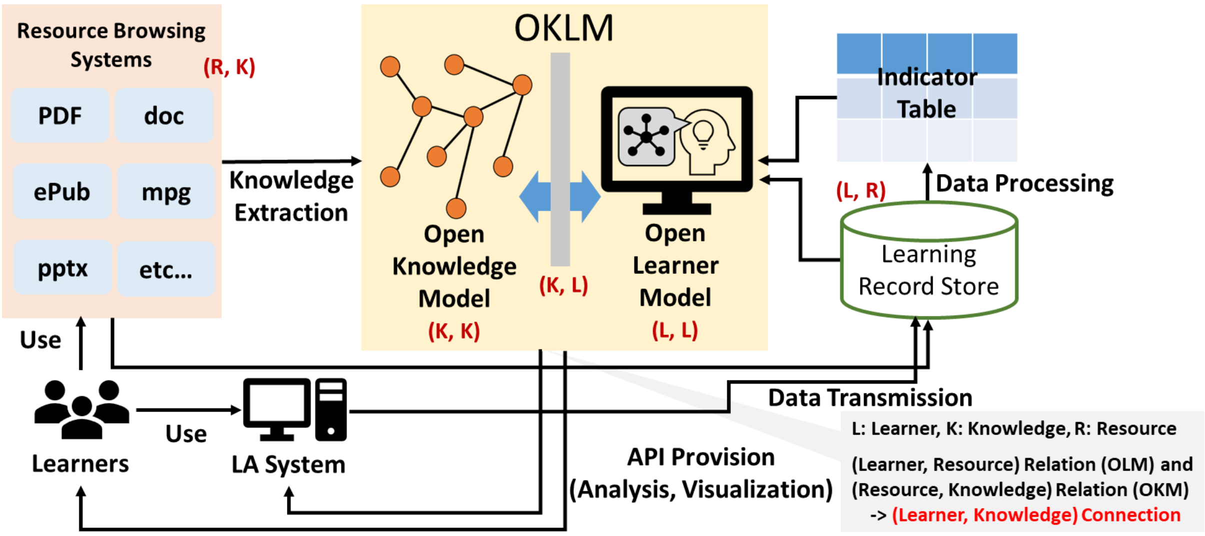



*Figure 1. Framework of Open Knowledge and Learner Model (OKLM)*

The essence of OKLM is to connect the learning activities conducted on the LA system with the knowledge information contained in various learning materials. First, knowledge is extracted from the learning resources using various methods, such as Flanagan et al. (2019). Second, learning logs collected from the LA system form a correspondence between learning activities and the resources they target. Using the above two types of relationships, it is possible to form a relationship between learning activities and knowledge items. In other words, estimating which knowledge is learned by a particular learning activity in each learning activity log is possible. These features ensure two types of openness: OKLM is knowledge domain-open (i.e., it works regardless of the knowledge domain of the material), and it can collect data regardless of the LA system used (i.e., system-open).

In addition, OKLM has a mechanism for publishing its own knowledge and learner models, as well as the results of visualization and analysis using the models, to learners and external LA systems that use OKLM. This is achieved through an API that OKLM exposes to external users and systems. This means that OKLM is also open to users and external systems for its models, a feature of OLM.

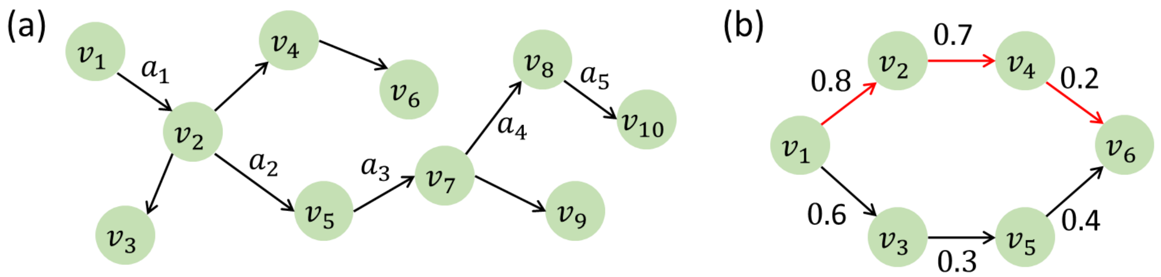
### *3.2 Comparison with Existing LMs*

There are two types of approaches to traditional learner modeling: overlay and buggy approaches (Nwana, 1990). Both learner modeling approaches need the process of knowledge modeling that comprehensively describes the knowledge to be learned, which often requires expertise and experience (Wang et al., 2017). On the other hand, OKLM builds knowledge models by extracting and reconstructing knowledge from learning materials using existing methods. Although the accuracy of the knowledge models in OKLM may be inferior to that of existing LMs, OKLM can guarantee the accuracy of the LM by using massive learning activity data and connect them with the knowledge models. This data-driven approach also guarantees the two kinds of openness of the OKLM, as mentioned in Section 3.1.

### 3.3 Mathematical Formulation of OKLM

#### 3.3.1 Preliminary: Definition of Knowledge Models

Human knowledge acquired through learning is not simply stored independently but generally represented as a network of semantically related knowledge items. In this study, the knowledge model is represented as a graph structure. First, let $G = (w, V, A)$ be a weighted directed graph of the knowledge model that represents the knowledge that the learner should target for learning. That is, $V$ is the set of knowledge items to be learned, $A \subseteq V \times V$ is a directed branch connecting knowledge items, and $w: A \longrightarrow [0,1]$ is a mapping to represent the weights of the branches. The weight of a branch is a numerical expression of how strongly the knowledge items connected by the branch are related to each other. In the following formulation, the larger the value of the weight $w(a)$ of a branch $a = (u, v)$, i.e., the stronger the relationship between knowledge items $u$ and $v$ from $u$ to $v$, the easier it is for a learner who knows knowledge item $u$ to analogize the meaning of $v$ from $u$.




$$p = a_1 a_2 a_3 a_4 a_5 \in P(v_1, v_{10})$$
$$w(p) = w(a_1) \cdot w(a_2) \cdot \cdots \cdot w(a_{10})$$

$$w_{opt}(v_1, v_6) = 0.8 \times 0.7 \times 0.2 = 0.112$$

*Figure 2.* (a) A pathway and its weight in the graph structure of a knowledge model. (b) The optimal value of the path weight for the knowledge analogy.

In the knowledge model $G$, which is a graph structure, the entire set of paths connecting two nodes $u, v$ is represented as $P(u, v)$ using the mapping $P: V \times V \to \bigcup_{n \in \mathbb{N}} A^n$. Then, the pathway $P \in P(u, v)$ can be expressed as $p = a_1 a_2 \cdots a_n$. The weight of the path $p$ is defined as $w(p) = \prod_{a \in p} w(a)$ (Figure 2a). Furthermore, the weight of the path connecting two nodes $u, v$ that has the maximum weight is denoted as $w_{opt}(u, v) = \max_{p \in P(u,v)} w(p) = \max_{p \in P(u,v)} \prod_{a \in p} w(a)$ (Figure 2b).

The following assertion follows from the above definition: $w(p)$ is the degree of analogy between $v$ and the two knowledge items $u, v$ connected by the path $p$, indicating how analogous it is to $v$ when the learner is familiar with $u$. Hence, $w_{opt}(u, v)$ is the optimal value for how analogous $v$ is to know $u$.

#### 3.3.2 Correspondence of xAPI logs and Learning Indicators

To understand the learning status of learners using the LA mechanism, it is necessary to estimate what aspects of knowledge each action in the learning support system using the LA method has led to learning. In this study, we consider mapping one of the six levels of Bloom's revised Taxonomy (Krathwohl, 2002) to the xAPI verbs, an international standard for representing learning behaviors (Figure 3).

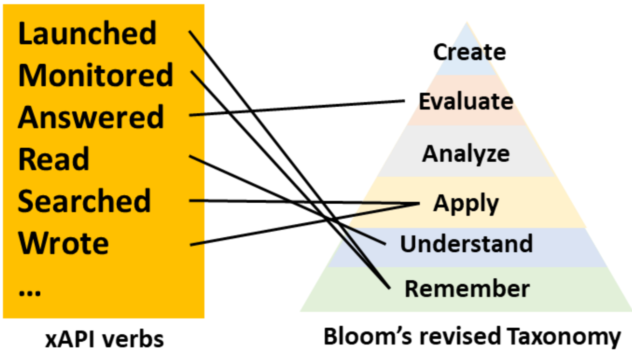



*Figure 3. Correspondence of xAPI verbs and the stages of Bloom's revised Taxonomy*

By converting the accumulation of learning activities corresponding to xAPI logs to the values of each indicator, these values can be denoted as follows: The value of learner $s$'s indicator $i$ for the knowledge item $v \in V$ at time $\tau$ is denoted as $x_i(s, v, \tau)$, where $i \in I = \{\text{Remember}, \text{Understand}, \text{Apply}, \text{Analyze}, \text{Evaluate}, \text{Create}\}$.

*3.3.3 Relative Proficiency*

Although we have formulated the learning behavior and the state of knowledge acquisition described above, treating this as "absolute learner proficiency" in the context of LA systems is very difficult. To treat the learner's proficiency with absolute criteria, it is necessary to guarantee qualitative aspects, such as the quality of the test to measure proficiency and the degree of effectiveness of learning behaviors in increasing proficiency. However, since LA guarantees data quality from the quantitative aspect of many learning behaviors and learners, it is impossible to treat learner proficiency with absolute criteria. Therefore, we define a group to which each learner belongs and treat the relative value of $x_i$ within that group as the relative proficiency. This can be formulated as follows.

First, assume that learner $s$ belongs to a specific group of learners $S$ (i.e., $s \in S$). At that time, we define **the relative proficiency** $rp_i(S, s, v, \tau) \in [0,1]$ of the indicator $i \in I$ for the knowledge item $v$ of the learner $s \in S$ within the group $S$ at time $\tau$ as follows:

$$rp_i(S, s, v, \tau) = \frac{1}{100} \max\left\{0, \min\left\{\frac{10\left(x_i(s, v, \tau) - \mu_i(S, v, \tau)\right)}{\sigma_i(S, v, \tau)} + 50, 100\right\}\right\} \tag{1}$$

However, $\mu_i(S, v, \tau) = \frac{1}{|S|}\sum_{s \in S} x_i(s, v, \tau)$ (the average of Indicator values in the population $S$), $\sigma_i(S, v, \tau) = \sqrt{\frac{1}{|S|}\sum_{s \in S}(x_i(s, v, \tau) - \mu_i(S, v, \tau))^2}$ (standard deviation of Indicator within population $S$). That is, the value (1) is the normalized deviation of the value of $x_i(s, v, \tau)$ within population $S$ within the range of 0 to 100 and divided by 100. Although deviations can theoretically take on all real values, it is reasonable to consider the values above 100 or below 0 as all outliers, given that these percentage is extremely low (about 0.000047%).

## 4. Application of OKLM to EFL Learning

In this section, we present a theoretical discussion of the possibilities of how the mathematical formulation of OKLM and the concept of relative proficiency, which we have introduced, can be used to support EFL learning and to represent the state of the learner. Here, we introduce a potential application of the fact that the relative proficiency $rp_i$ is a function of the learner, the population to which the learner belongs, the knowledge item, and time.

### *4.1 Knowledge Analogy: Mathematical Representation of Spreading Activation Model*

### *4.1.1 Modeling the Easiness of Learning by Analogy*

The concept of relative proficiency presented in this paper can be used to model the analogy of knowledge in learning activities mathematically. The analogy of knowledge is using already known information to infer the meaning of knowledge about an unknown matter (Citation). Here is an example in the knowledge of EFL learning: Suppose a student does not know the meaning of the word “cooperation.” If the student knew that the prefix “co-“ means “together” and the word “operation” means “manipulation” or “action,” then by analogy with this knowledge, s/he could assume that “cooperation” means “to work together.”

Using the relative proficiency (1), we can define **the easiness of learning** knowledge items $el_i(S,s,v,\tau) \in [0,1]$ as follows:

$$el_i(S,s,v,\tau) = \max_{u\in V} w_{opt}(u,v) rp_i(S,s,u,\tau) \quad (2)$$

This formula shows that when the learner $s$ has been proficient in the knowledge item s/he had learned before (denoted as $u$), s/he can learn a new knowledge item $v$ related to $u$ with the ease of level (2). This phenomenon, in which learning one knowledge item makes it easier to learn another related knowledge item, originates from the analogy of knowledge in learning.

This formula can be interpreted as a mathematical representation of the spreading activation model of knowledge learning. The spreading activation model, a finding in the field of cognitive psychology, is based on the premise that concepts acquired by humans are stored as a network structure in the brain and that when a concept is recalled, related concepts are also activated, thereby promoting the use of the concept (Collins & Loftus, 1975) (Figure 4). In other words, based on the activation-diffusion model of knowledge, $rp_i(S,s,v,\tau)$ indicates the degree of knowledge mastery. In contrast, $el_i(S,s,v,\tau)$ indicates the degree of knowledge activation.

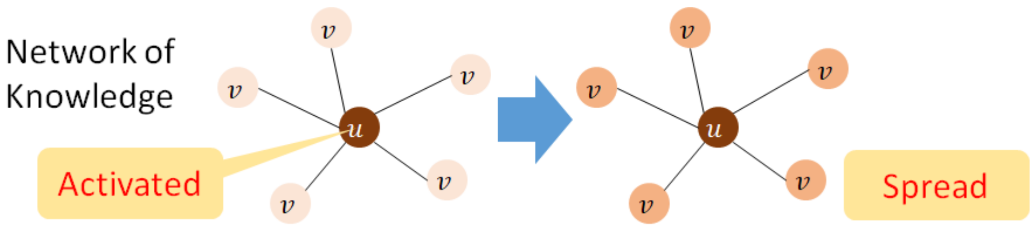



*Figure 4. The spreading activation model of knowledge*

### *4.1.2 Modeling Comprehensive Proficiency*

As shown above, knowledge with a high degree of knowledge activation can be learned more quickly than knowledge items with low activation once the opportunity arises. In other words, the ease of knowledge learning derived from knowledge analogy can be considered the learner's potential proficiency level. If we equate this with the explicit proficiency of knowledge once learned, we can define **the comprehensive proficiency** of knowledge as follows:

$$cp_i(S,s,v,\tau) = c_{cp1} rp_i(S,s,v,\tau) + c_{cp2} el_i(S,s,v,\tau) \quad (3)$$

The $c_{cp1}$ and $c_{cp2}$ are constants, $c_{cp1} \geq 0, c_{cp2} \geq 0,$ and $c_{cp1} + c_{cp2} = 1$.The constants $c_{cp1}$ and $c_{cp2}$ are values that express whether the learner's apparent proficiency or the easiness of learning is more important. The larger $c_{cp2}$ is, the more importance is placed on the learner's easiness of learning specific knowledge, i.e., the potential proficiency introduced by the analogy of known knowledge.

## *4.2 Representation of Learning Materials*

### *4.2.1 Modeling Learning Materials*

In the same way that we defined relative proficiency, we can also use mathematical formulas to describe the learning objectives the learner must understand. To understand the contents of EFL texts or materials, one must be proficient in the required knowledge (i.e., vocabulary, grammar, etc.) to the required depth. This is expressed as follows: For a learning material $m$, consider the set of knowledge items $V_m$ required to understand $m$. Let $up_i(S,m,v,\tau)$ denote the relative proficiency of the knowledge item $v \in V_m$ at time $\tau$ in terms of Indicator $i \in I$ that is required to understand the material $m$. In this case, the state in which the material $m$ is understandable, i.e., the learner has mastered all knowledge in $V_m$ to a certain level, can be expressed by the equation (4):

$$\forall v \in V_m, up_i(S,m,v,\tau) \leq cp_i(S,s,v,\tau) \qquad (4)$$

If the equation is satisfied, we can predict that the learner will be able to understand the material, which may help predict the learner's performance and learning behavior in the future.

Although the above discussion is based on materials, it can be applied not only to materials but also to the mastery levels required to solve learning tasks. In other words, it is sufficient to define the knowledge items required to achieve a certain goal and their level of mastery. For example, it may be possible to predict in advance whether a learner will be able to solve a certain task. It may also be possible to predict whether a learner will pass or fail an exam before s/he takes it.

### *4.2.2 Material Recommendation*

Next, we introduce the possibility of using the concepts derived from relative proficiency introduced above to recommend learning materials. In the context of e-learning, the number of digital material repositories continues to grow, which means that learners face an excessive number of online learning resources (Ochoa & Duval, 2009). Therefore, creating a learning environment in which learners can select appropriate learning materials using a material recommendation system is essential. In this paper, we take a simple approach of expressing how much a specific material should be recommended to a particular learner in terms of a score and then strongly recommending the material in the order of its higher score.

There are various possible methods for scoring learning materials. The first method is to consider the learner's proficiency level in the knowledge items included in the material. Materials that contain much knowledge that a learner has already mastered well are accessible for the learner to work with and should be recommended (Condition 1). On the other hand, materials that contain much knowledge that a learner has not mastered well before should also be recommended because learning these materials will broaden the learner's knowledge and help the learner overcome his/her difficulties (Condition 2). In this case, we can use the value of the function $cp_i$ mentioned above as an index for scoring the knowledge contained in the materials. In other words, a new function $f_i^{cp}$ in the equation (5) can be defined using (3) as follows and used as a basis for recommending educational materials. In other words, the learners rate the knowledge in a particular material they have mastered in Condition 1 highly, while they rate the knowledge, they have not mastered in Condition 2 highly.

$$f_i^{cp}(S,s,v,\tau) = \begin{cases} cp_i(S,s,v,\tau) & (\text{in Condition 1.}) \\ 1 - cp_i(S,s,v,\tau) & (\text{in Condition 2.}) \end{cases} \qquad (5)$$

The following method considers the time elapsed since the learner last learned a particular knowledge item. Once learned knowledge is not reviewed, it will likely be forgotten if time passes without review. Therefore, it is meaningful for a learning support system to encourage learners not to forget the knowledge they have learned. Therefore, the function $f_i^{forg}$ in the equation (6) is defined in this paper to recommend learning materials, which can be highly evaluated if the elapsed time since the last learning of a specific knowledge is extended. In the equation (6), $\tau_{last}^{i,s,v}$ refers to the time when the learner $s$ last took a learning action for a knowledge item $v$ before the time $\tau$ at a certain index $i \in I$. $c_g$ is a constant, and $c_g > 0$.

$$f_i^{forg}(s,v,\tau) = 1 - \frac{1}{c_g\left(\tau - \tau_{last}^{i,s,v}\right) - 1} \tag{6}$$

The third method is to give high ratings to knowledge items that are well mastered by other learners. In this paper, we assume that learners are learning in a group of many learners. In such an environment, it is highly probable that the knowledge that other learners have learned well will be regarded as important in learning the material and the domain. Therefore, we propose a recommendation method that gives high ratings to materials that contain a large amount of knowledge that is well studied by other learners. This can be calculated by the function $f_i^{sig}$ in (7).

$$f_i^{sig}(S,s,v,\tau) = \frac{1}{|S|-1} \sum_{\substack{s' \in S \\ s' \neq s}} rp_i(S,s',v,\tau) \tag{7}$$

Synthesizing all the above methods, we can define the recommendation score of a teaching material $m$ in an indicator $i \in I$ for a learner $s$ belonging to a group $S$ as (8), where $\alpha \geq 0, \beta \geq 0, \gamma \geq 0$, and $\alpha + \beta + \gamma = 1$. By adjusting the values of these constants $\alpha$, $\beta$, and $\gamma$, the system can adjust which of the above three perspectives is more important.

$$\begin{aligned} &\mathrm{score}_i(S,s,m,\tau) \\ &= \frac{1}{|V_m||V|} \sum_{u \in V_m} \sum_{v \in V} w_{opt}(u,v) \left(\alpha f_i^{cp}(S,s,u,\tau) + \beta f_i^{forg}(s,u,\tau) + \gamma f_i^{sig}(S,s,u,\tau)\right) \end{aligned} \tag{8}$$

Like the modeling of task and instructional materials, this recommendation method can also be applied to the recommendation of tasks and learning objectives. Since they all encompass specific knowledge items, recommendations can be generated for tasks and learning objectives in a similar manner, provided that the learner's past learning activities are defined for those items.

## 5. OKLM Use Case and Demonstration Plan

### *5.1 The Use Case of OKLM in an EFL Learning Activity*

To show that OKLM can be used in a variety of learning contexts, it is worthwhile to provide some examples of its use. In this section, we present an example of OKLM use in EFL reading and writing activities using e-book materials.

Let us suppose that a learner is reading an EFL material with an e-book. This EFL material is registered as a PDF file to the e-book reader beforehand, and the vocabulary and grammar included in the content are extracted and reconstructed as knowledge models at that time. When learners are exposed to English vocabulary and grammar in the course materials, the "Understand" and "Remember" indicators in the OKLM increase for the knowledge items they are exposed to. In this way, the learner's knowledge is roughly modeled by massive educational big data, and data on which knowledge items are grasped and from which perspective and to what extent are accumulated. Next, suppose that the learner uses the vocabulary and grammatical knowledge learned in the e-book for writing and speaking. When the LA system detects such a use of knowledge, the value of the "Apply" indicator for the knowledge is increased. Thus, OKLM builds a data-driven model of the learner's knowledge state by collecting data on how proficient the learner is in each knowledge item and from which perspective. Of course, the constructed models can be viewed by users including learners and teachers. Moreover, OKLM not only addresses each learning activity, but also bridges each of them.

### *5.2 Demonstration Plan for Verification of OKLM*

A plan for empirical validation of the OKLM framework is threefold: (a) comparison with a simple LM, (b) investigation of the efficiency of learner modeling in OKLM, and (c) validation of the effectiveness of learning support using OKLM. In (a), we aim to demonstrate that the OKLM's unique LA-based structure, which is not found in traditional OLMs, is effective in learning support. This part includes an experiment using a learning support system using a simple LM to investigate its learning and motivational effects, as well as the level of system acceptance. Part (b) is intended to demonstrate the effectiveness of OKLM's learner modeling methodology for representing learner states. This involves using the OKLM-based learning system and comparing the learner's learning outcomes over the period of use with the features in the OKLM data. Part (c) will measure the effectiveness of the OKLM's three kinds of "openness" mentioned in Section 3.1: namely, openness for knowledge domains, systems, and users. This part needs knowledge and learner modeling in different domains, modeling using different learning support systems, and verification of the effect on understanding learners' states, motivation, or characteristics by users, respectively.

## 6. Limitations

The reliability of relative proficiency significantly depends on how indicators are calculated from the LA-based learning support system. Specifically, the reliability of the formula (1) varies based on how $x_i(s, v, \tau)$ is derived from the accumulated learning logs, affecting the effectiveness of learning support. Since the calculation method of $x_i$ may differ across LA systems using OKLM, a universally valid method is challenging to establish. However, once calculable, the framework's usefulness can be verified. Although this paper is theoretical, empirical verification with actual learning logs is urgently needed.

Additionally, while the formulation addresses the proficiency of individual knowledge items, it does not fully consider the proficiency of connections among these items. In EFL learning, for instance, semantic links between vocabulary or grammar learning paths should be factored in. Learners may have varying proficiency levels not only in individual knowledge items but also in how they perceive connections between them. This paper's knowledge model assumes uniform connections for all learners, limiting its ability to offer personalized learning support. Enhancing the model's expressiveness requires considering proficiency in these connections.

## 7. Conclusion and Future Work

In this study, we present a mathematical formulation of OKLM developed to understand learners' states in the LA context. We also propose calculating the "relative proficiency" of a group of learners concerning a particular knowledge based on their learning activity logs accumulated through their use of the LA system. Some examples of possible applications of this mathematical formulation to learning support are also presented. For example, it was shown that analogy of vocabulary knowledge, modeling learning materials, and material recommendations can be represented by the formulation.

Since only theoretical considerations and the formulation of the OKLM were made in this paper, it is an urgent task to verify the effectiveness of this model using actual data from the learning activity logs stored in the LA system. It is also an essential issue for the future to improve the model based on the problems found in this validation process. Second, while this paper introduced some examples of learning supports for EFL learning, OKLM has the potential to be introduced into another broader context, such as learning math. We should broaden the area of the application of OKLM to various contexts other than EFL. Furthermore, to make OKLM more useful as an LA system, it is necessary to establish a feedback mechanism for various stakeholders involved in education, such as learners, their parents, and teachers. This research contributes to the development of OKLM not only to support learning by predicting learner performance and detecting at-risk students, but also to open new possibilities for LA, such as recommending teaching materials and modeling of learners.

## Acknowledgements

This work was supported by CSTI SIP Program Grant Number JPJ012347, NEDO JPNP20006, and JST Grant Number JPMJFS2123.

## References

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018, March). Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41-50).

Bull, S. (2020). There are open learner models about!. *IEEE Transactions on Learning Technologies, 13*(2), 425-448.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, *4*, 253-278.

Falmagne, J. C., & Doignon, J. P. (2010). *Learning spaces: Interdisciplinary applied mathematics*. Springer Science & Business Media.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5-6), 304-317.

Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge map creation for modeling learning behaviors in digital learning environments. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)* (pp. 428-436). Society for Learning Analytics Research (SoLAR).

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hsu, M. H. (2008). A personalized English learning recommender system for ESL students. *Expert Systems with Applications*, *34*(1), 683-688.

Huang, Y., & Zhu, J. (2021). A personalized English learning material recommendation system based on knowledge graph. *International Journal of Emerging Technologies in Learning (Online)*, *16*(11), 160.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218.

Nwana, H. S. (1990). Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, *4*(4), 251-277.

Ochoa, X., & Duval, E. (2009). Quantitative analysis of learning object repositories. *IEEE Transactions on Learning Technologies*, *2*(3), 226-238.

Takii, K., Flanagan, B., Li, H., Yang, Y., & Ogata, H. (2022, November). Explainable English Material Recommendation Using an Information Retrieval Technique for EFL Learning. In *The30th International Conference on Computers in Education (ICCE2022)* (pp. 561-570).

Takii, K., Koike, K., Horikoshi, I., Flanagan, B., & Ogata, H. (2023). OKLM: A universal learner model integrating everyday learning activities with knowledge maps. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)* (pp. 191-193). Society for Learning Analytics Research (SoLAR).

Wang, J., Flanagan, B., & Ogata, H. (2017, July). Semi-automatic construction of ontology based on data mining technique. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 511-515). IEEE.

Zabolotskikh, A., Zabolotskikh, A., Dugina, T., & Tavberidze, D. (2021). Creating individual learning paths in the Moodle plugin for undergraduate students to study English grammar. *Education and Information Technologies*, *26*(1), 617-637.