

Doctoral Dissertation

OKLM: Open Knowledge and Learner Model for Learning Analytics

Kensuke TAKII

Supervisor: Professor Hiroaki OGATA

October 2025

Department of Social Informatics
Graduate School of Informatics
Kyoto University

Abstract

The use of big data in education has garnered significant attention in recent years, and Learning Analytics (LA) is being researched globally as a data-driven methodology for understanding and optimizing human learning. However, while existing LA richly captures learners’ behaviors, it rarely represents what knowledge is being learned and to what extent it is understood. To overcome this gap, this thesis proposes the Open Knowledge and Learner Model (OKLM), which handles not only learner behavior but also acquired knowledge and thereby enables LA systems to track “what was learned” at a content level. OKLM is designed around three kinds of openness: knowledge-domain openness, learner openness, and LA system openness. These openness principles enable the models generalize across domains, remain interpretable to learners and teachers, and interoperate with heterogeneous LA platforms.

Methodologically, this thesis formalizes OKLM with (i) a graph-based knowledge model and mappings from xAPI logs to learning indicators, and (ii) Information Theory on Learning Analytics and Knowledge (ITLAK), which quantifies the information content (novelty/uncertainty) of each knowledge encounter. Together, these allow relative proficiency to be estimated over knowledge graphs and make each learning action analyzable as information bearing on specific knowledge components.

OKLM is implemented on the Learning and Evidence Analysis Framework (LEAF) and validated through three empirical studies: (1) explainable extensive reading recommendations using vocabulary knowledge, (2) grammar learning with information-theoretic analysis of short- and long-term effects, and (3) academic paper recommendations that operationalize comprehensive proficiency. Across studies, OKLM enables interpretable, knowledge-aware recommendations; ITLAK reveals a complementary pattern that involves novelty and repetition; and new indicators (knowledge coverage, mean proficiency, information content) offer actionable evaluation beyond scores or time-on-task.

The contributions of this thesis are threefold: (1) proposing, designing, developing, and implementing OKLM that satisfies the three openness requirements; (2) grounding a Knowledge-Aware LA (KALA) paradigm for LA through OKLM and ITLAK; and (3) demonstrating objective, quantitative representation of learners’ knowledge from everyday logs, thereby advancing LA from behavior-centric monitoring toward interpretable, knowledge-centric support for learning and teaching.

Contents

1	Introduction	3
1.1	Research Background	3
1.2	Problems and Solutions	4
1.2.1	The Absence of Knowledge in Educational Data	4
1.2.2	Lifelong Learning Support	5
1.2.3	OKLM as the Solutions	5
1.3	Contributions	6
1.4	Research Questions	6
2	Literature Review	9
2.1	Traditional LA and Its Limitations	9
2.1.1	Definitions of LA and Their Implications	9
2.1.2	Limitations of LA's Data-Drivenness	10
2.2	Knowledge Modeling	10
2.3	Learner Knowledge Representation and LA	11
2.4	Implications for OKLM	12
3	Design of OKLM	15
3.1	The OKLM Framework Concept	15
3.2	System Requirements	15
3.2.1	Knowledge-Domain Openness	16
3.2.2	Learner Openness	16
3.2.3	LA System Openness	16
3.3	Mathemtaical Formulation	16
3.3.1	Formulation by Relative Proficiency	17
3.3.2	Formulation by ITLAK	19

3.4	OKLM Visualization	21
4	System Implementation	25
4.1	LEAF System	25
4.2	System Components	27
5	Experimental Studies	31
5.1	Study 1: Knowledge Modeling & Simple Learner Modeling	31
5.1.1	Research Proposals and Research Questions	31
5.1.2	Recommendation Platform	32
5.1.3	Recommendation Mechanism	34
5.1.4	Sub-Experiment for Automated Material Difficulty Estimation . . .	39
5.1.5	Setting	40
5.1.6	Result	42
5.1.7	Discussion	49
5.2	Study 2: Learner Modeling	52
5.2.1	Research Proposals and Research Questions	52
5.2.2	Experimental Setting	54
5.2.3	Analysis Method	55
5.2.4	Result	58
5.2.5	Discussion	61
5.3	Study 3: OKLM Application to Learning Support	67
5.3.1	Research Proposals and Research Questions	67
5.3.2	Paper Recommendation	68
5.3.3	Setting	70
5.3.4	Data Collection	73
5.3.5	Analysis Method	75
5.3.6	Result	78
5.3.7	Discussion and Implications	83
5.4	Summary of the Three Studies	89
5.4.1	Satisfaction of Requirements in Each Study	89
6	General Discussion	91
6.1	Theoretical Implications	91

6.1.1	Reconceptualizing Learner Modeling in LA	91
6.1.2	Formalizing Information in Learning	92
6.1.3	A Network-Based Knowledge Perspective	92
6.1.4	Toward a Unified Theory of KALA	93
6.2	Pedagogical Implications	93
6.2.1	Explainable Recommendations for Learners and Teachers	93
6.2.2	Novelty and Repetition as a Design Pattern	94
6.2.3	Classroom Management and Formative Assessment	94
6.2.4	Evaluation Indicators for Educational Effectiveness	94
6.2.5	Scalability and Cross-Context Application	95
6.2.6	Toward Evidence-Informed Educational Practice	95
6.3	Limitations and Future Work	95
7	Conclusions	97
	References	101
	Appendices	107

List of Figures

1.1	Overview of the research approach	8
3.1	(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 (Takii, Liang, et al., 2024).	17
3.2	Correspondence of xAPI verbs and the stages of Bloom’s revised Taxonomy (Takii, Liang, et al., 2024).	18
3.3	An OKLM structure and an example of calculating information content of ITLAK (Takii et al., 2025). Learning logs retrieved from an LA system are linked to knowledge elements, as shown in the table in this figure. The fewer times an element has been encountered, the higher its information content. In this study, we calculate using base 2. Although α is set as 0 in these examples, we guarantee $\alpha > 0$ in an actual setting.	20
3.4	Visualization of OKLM. Green and yellow nodes show knowledge compo- nents learners are familiar and unfamiliar with, respectively.	23
4.1	Structure of the LEAF system	25
4.2	An example xAPI statement	27
4.3	Framework of OKLM	27
4.4	Sample of OKLM internal data. Visualization utilizes Neo4j function. . . .	28
5.1	An overview of the recommendation platform	33
5.2	UI of the picture book recommender system implemented in the GOAL system	34
5.3	Numbers of daily accesses and clicks on the recommendation page	44
5.4	Explained variance for the clustering. Based on this, we decided on three clusters.	48

5.5	The breakdown of answers and learners' activities by each cluster	48
5.6	The spreading activation model of knowledge	68
5.7	Screenshot of a paper recommendation system. The higher up the list, the more strongly recommended the paper is.	71
5.8	(a) Number of recommendation displays per task, (b) Number of clicks on recommended papers across the three tasks, (c) Acceptance rate (clicks/displays) per task, and (d) Number of papers selected before final decision.	80
5.9	Violin plots of students' responses to subjective evaluation items (1-5 Likert scale)	81
5.10	Comparison of report-based learning outcomes across three recommenda- tion conditions. (a) Report score, (b) Total word count, and (c) Vocabulary size	82
5.11	Distribution of responses to three subjective evaluation items	83

List of Tables

5.1	Sentences that explain why the materials were recommended. They depend on the difference between the difficulty of the material and the learner's difficulty preference.	37
5.2	Results of an evaluation of an equation that computes the difficulty of books	39
5.3	Summary of participants and experimental settings	41
5.4	The number of picture books classified by CEFR levels	41
5.5	List of questions in the post-poll and their class in TAM. Each question item was rated on a five-point Likert scale.	43
5.6	Daily means of the number of accesses and clicks	44
5.7	Acceptance rate of the recommendation. The value was calculated by access to the recommendation or the number of clicks on the title(s) of the recommended books.	45
5.8	Summary of the post poll results. This shows the number of participants who answered each question as 1 (strongly disagree) to 5 (strongly agree) and the mean values of the answers for each question.	46
5.9	Correlation between learners' activities and poll answers	46
5.10	The clustering result using the poll answers (mean (std.)). It also shows the breankdown of answers and learners' activities by cluster.	47
5.11	Notation Used in This Study	56
5.12	Parameter settings and algorithmic focus in each recommendation condition	72
5.13	Correlation coefficients between OKLM indicators and learning outcomes .	79
5.14	Requirement fulfillment of OKLM's three openesses	89
1	Complete list of the post-questionnaire items.	108

Acknowledgement

Above all, I would like to sincerely thank my supervisor, Prof. Hiroaki Ogata, for his support and guidance throughout the past six years, from my admission to the master's program to my current work in the doctoral program. As a student who knew little about anything, I received invaluable opportunities to conduct research in a cutting-edge environment, exposure to the world's most advanced academic knowledge, and exceptionally thorough guidance on how to carry out research and shape results as a researcher. I firmly believe that without his guidance, I would not be here today, conducting research in the doctoral program and writing my thesis.

I would also like to thank my advisers: Dr. Donghui Lin, Prof. Hisashi Kashima, and Prof. Takayuki Ito for their critical opinions and insightful comments, which provided me with research suggestions and opportunities for growth as a researcher. They made time for discussions amidst their busy schedules to support my research, and they provided me with opportunities to refine my research concepts by examining them from various perspectives beyond the context of understanding and supporting learning and education. Their advice was instrumental in helping me establish my research as a cohesive design, and it also equipped me with methodologies to re-examine my work from different viewpoints.

I am also grateful to the laboratory members who reviewed and guided my research: Dr. Brendan Flanagan, Dr. Kento Koike, and Dr. Changhao Liang. They offered me countless insights essential for becoming an independent researcher. Beyond research methodologies and techniques, I learned immeasurably from them about research philosophy and the importance of refining it, as well as the joys and resolve required to live as a researcher. Their teachings remain ingrained in my very being. Thank you so very much for helping me grow as a researcher.

Furthermore, I would like to express my gratitude to the members of my research lab who have consistently supported me in various ways. Without you all, I could never have endured the demanding yet rewarding life in the Ogata Laboratory. Having played, debated, researched, and lived alongside you is a source of pride in my life. Thank you for these irreplaceable memories.

Finally, I express my deepest gratitude to my parents, Junichi and Nobuko, for their strong support, encouragement, advice, help, and delicious meals. I also offer my gratitude

to my little sister, Mayu, for giving moments of peace through lighthearted conversation. I truly would not be here without you. I would also like to express my gratitude to all my friends in the club: our discussions about music have enriched my life. Thank you all very much for being there for me.

Chapter 1

Introduction

1.1 Research Background

The penetration of information technology into our daily lives has been advancing at an increasingly rapid pace in recent years. The field of human learning and education is no exception. Digital learning and educational tools are increasingly enhancing the efficiency of learning and education, transforming education to be more accessible, interactive, and data-driven. Notably, in Japan, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has proposed the GIGA School Initiative, which aims to provide every child and student nationwide with a personal computer and network access (MEXT, 2020).

Notably, the shift toward data-driven learning and education has enabled the capture of human learning processes at behavioral level. Learning Analytics (LA) is a promising research field that analyzes the educational data accumulated in this manner, aiming to understand and optimize learning and its contextual occurrence. LA has enabled the acquisition of various insights and learning support that were difficult with existing approaches: such as predicting academic performance (Jayaprakash et al., 2014), capturing learning patterns (Kizilcec et al., 2013), and designing dashboards that utilize these insights (Papamitsiou & Economides, 2014). This data-driven approach has given rise to the paradigm of “educational data science,” which uses learning and educational data to understand learning. Furthermore, from an engineering perspective, LA has significantly contributed to improving learning and education in educational settings. Its results are made available to stakeholders such as learners and educators through the LA Dashboard (LAD) (Verbert et al., 2013).

1.2 Problems and Solutions

1.2.1 The Absence of Knowledge in Educational Data

LA primarily analyzes learning action units—that is, “what learning actions were taken”—but due to its data-driven nature, it has struggled to handle phenomena that cannot be expressed through data. The prime example of this is knowledge. In LA, capturing learners’ knowledge—that is, “what and how much they learned”—was fundamentally difficult.

Attempts to capture learners’ knowledge using computers predate LA. This research was particularly prevalent in learner models within existing intelligent tutoring systems (Nwana, 1990). However, these studies were insufficient in the following respects:

- 1. A predefined and elaborate knowledge model is required.** Many previous learner models are based on the premise of building models specialized for a particular learning domain, which required preparing predefined and well-developed knowledge models. For example, learner models have been developed to support learning in the mechanics field of physics (Aikawa et al., 2024) or to support learning specific programming languages. Generally, constructing accurate knowledge models requires specialized expertise and manual judgment, making it costly and time-consuming (Khadir et al., 2021). Furthermore, due to this field dependency, building a universal learner model was extremely difficult.

- 2. The granularity of observable learning behaviors is coarse.** Many current learning support systems capture only coarse-grained learning behaviors, such as whether an individual response to a question is correct or incorrect (Nakamoto, 2025). Consequently, compared to systems utilizing LA technology, they have a smaller volume and lower density of behaviors that can serve as the basis for learner modeling.

In other words, we need a method for constructing learner models that minimizes the effort required to build knowledge models while still capturing learner behavior at a detailed granularity. Therefore, the problem we must solve—and for which this thesis provides a solution—is “Can we handle knowledge acquisition using LA methodologies?”

1.2.2 Lifelong Learning Support

Another area posing challenges in LA-based learning and educational support is lifelong learning. Lifelong learning refers to the continuous development of knowledge and skills across formal, non-formal, and informal contexts throughout an individual’s life (UNESCO, 2015). In the field of LA, supporting lifelong learning has become a pressing issue, since the learner’s digital traces are usually fragmented across multiple institutions and systems. To address this challenge, Ocheja et al. (2019) proposed the Blockchain of Learning Logs (BOLL) framework, which enables learners to securely transfer and verify their learning records among institutions by storing hashed learning logs and access permissions on a blockchain. BOLL thereby seeks to provide continuity of learning data and overcome the “cold-start problem” often faced by LA platforms when new learners enter unfamiliar environments.

Despite such progress, the computational representation of learners’ knowledge states remains underdeveloped in lifelong-learning-oriented LA. One critical reason is that learner models are typically embedded within individual LA systems or Learning Record Store (LRS), making it difficult to inherit or reuse a learner’s model when they move between schools or learning environments. As a result, each system must reconstruct the learner model from scratch, limiting the continuity and interpretability of accumulated knowledge. Therefore, we raise another problem to solve: “Can a learner’s knowledge be modeled in a form that remains interpretable and inheritable across different LA platforms?”

1.2.3 OKLM as the Solutions

To address these above fundamental issues, we propose the Open Knowledge and Learner Model (OKLM), a learner modeling framework based on the LA methodology. As its name suggests, OKLM possesses three types of “Openness” as its properties: domain-knowledge openness, learner openness, and LA system openness (see Section 3.2). These properties are derived from the characteristics of existing open learner models and LA methodologies, equipping OKLM with the means to solve the problems presented earlier. This thesis demonstrates that the design of OKLM is practically feasible and describes its empirical validation of effective functioning in educational settings.

1.3 Contributions

In this research, we propose, design, develop, and implement OKLM to solve the above problems. Our research contributions are threefold:

1. Proposed, designed, developed, and implemented a learner model framework that satisfies the three openness criteria: knowledge-domain openness, learner openness, and LA system openness.
2. Implemented the principle of Knowledge-Aware LA (KALA) through the learner model.
3. Proposed and demonstrated a method for objectively and quantitatively expressing the learning of knowledge components in learner modeling using LA.

1.4 Research Questions

This thesis focuses on answering the following research questions related to knowledge modeling (RQ1), learner modeling (RQ2), and application to learning support (RQ3). These research questions also refer to the usage of these models:

RQ1 (Pub. 1): How can OKLM extract knowledge components from learning materials and model them?

RQ2 (Pub. 2): How can OKLM detect knowledge comprehension levels from learners' actions and model them?

RQ3 (Pub. 3): How effectively can OKLM support teaching and learning in other LA applications?

In chapter 1, we share the background of this research and the challenges to be addressed, present OKLM as the solution, and outline the research questions for this study. Chapter 2 discusses what LA fundamentally constitutes as an academic field and the issues stemming from its data-driven methodology. It also describes research trends regarding prior studies on representing learner knowledge and their relationship to LA.

Chapter 3 discusses the design of OKLM. First, we provide an overview of the OKLM framework, followed by an explanation of the “three types of openness,” which are also system requirements. Subsequently, we detail relative proficiency and Information Theory

on Learning Analytics and Knowledge (ITLAK) as the mathematical descriptions of how OKLM represents learner knowledge.

Chapter 4 describes the implementation of OKLM. First, it explains the Learning and Evidence Analytics Framework (LEAF) system, the underlying LA framework relied upon by OKLM in this research and demonstration. Subsequently, it describes OKLM's system components, their interconnections, and their respective functions.

Chapter 5 constitutes the main part of this thesis. This chapter describes three types of empirical experiments concerning the implementation of OKLM and the verification of its effectiveness. Each experiment corresponds to Research Questions 1 through 3. The first experiment utilizes a material recommendation system employing English extensive reading learning materials. The knowledge representation of the materials and learners' English vocabulary that this recommendation system relies upon forms the foundation of OKLM. The second experiment focuses on intensive English reading learning for junior high school students, modeling learners' knowledge state regarding English grammar. The third experiment utilizes a paper recommendation system in an undergraduate English literature reading course, where OKLM is first concretely established and applied to learning support. Here, for the first time, the learner's "level of knowledge" is modeled using a theoretical approach independent of simple exposure frequency. Finally, we summarize the results of the three empirical experiments.

Chapter 6 examines the implications of this study on OKLM from both theoretical and practical perspectives. It also discusses the limitations of this research and future research directions. Finally, Chapter 7 presents the conclusions of this study.

Figure 1.1 provides an overview of the structure of this thesis. It summarizes the motivation for this research, research gaps, and the correspondence between each RQ and the empirical experiments.

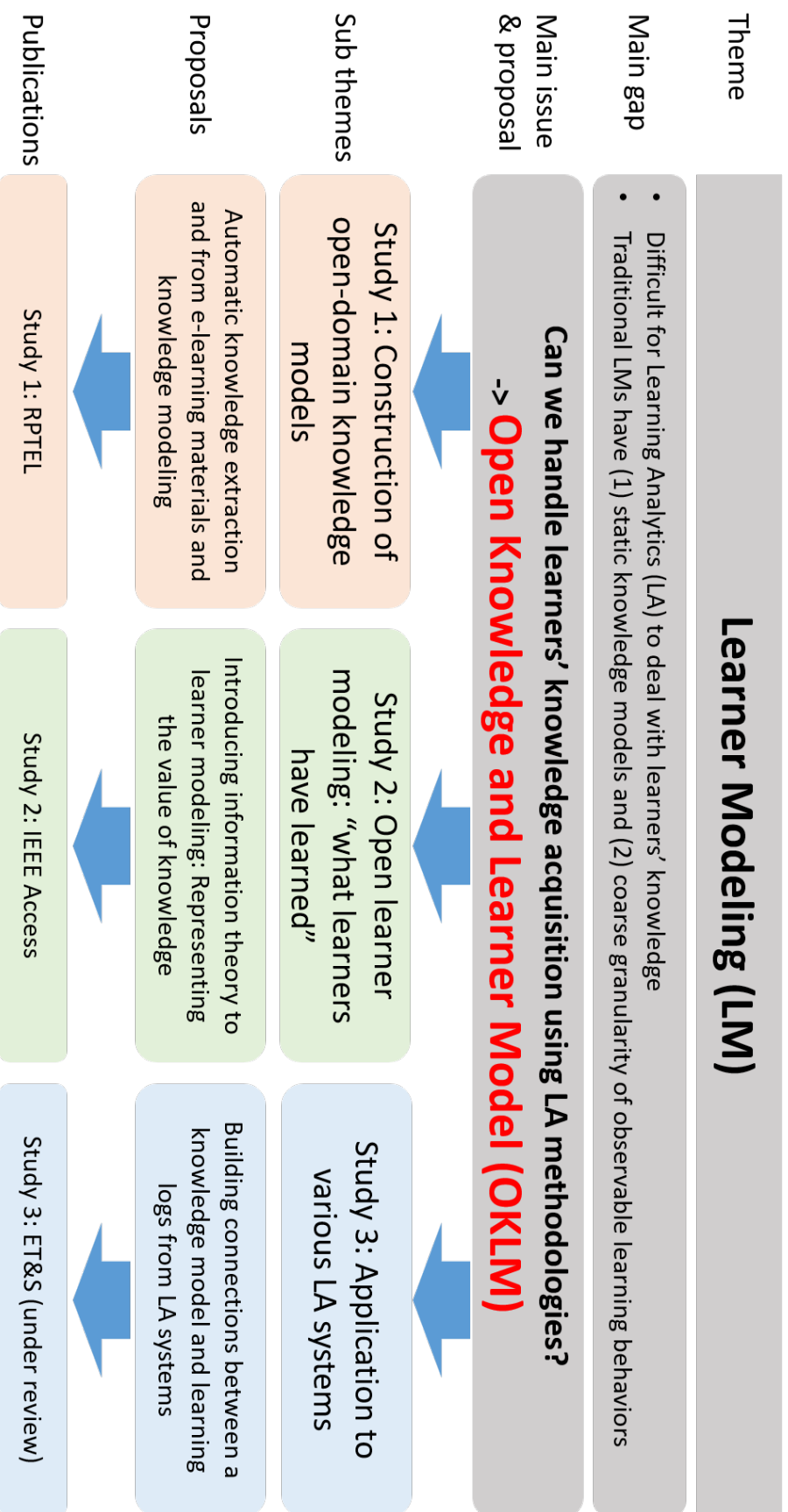


Figure 1.1: Overview of the research approach

Chapter 2

Literature Review

2.1 Traditional LA and Its Limitations

2.1.1 Definitions of LA and Their Implications

The field of LA has developed rapidly since the inaugural International Conference on Learning Analytics and Knowledge in 2011, held by Society for Learning Analytics Research (SoLAR). Traditionally, the most well-known and widely cited definition is by Ferguson (2012), “the measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.” This definition encompasses both understanding and optimization of learning, but by emphasizing “understanding,” it positions LA as a data-driven methodological endeavor within the science of learning and education. Furthermore, in the first edition of the Handbook of Learning Analytics (Knight & Buckingham Shum, 2017), LA is described as “a new form of assessment instrument” that utilizes data, designed and used based on some epistemology, pedagogy, and assessment.

The revised handbook (Lang et al., 2022), updated as the field matured, reexamines the definition of LA and proposes understanding it from four perspectives: concern, opportunity, field of inquiry, and community. Specifically, LA is described as an academic field that, against the backdrop of educational data as a concern and opportunity, pursues the understanding and support of learning through collaboration between humans and opportunities. It is depicted as an activity formed and sustained by communities. Furthermore, according to the latest definition by SoLAR (2025), LA is defined as “the collection, analysis, interpretation and communication of data about learners and their learning that provides theoretically relevant and actionable insights to enhance learning and teaching.”

By tracing these transitions, we can discern how the meaning of the endeavor known as LA has evolved. Specifically, LA has transformed from an engineering endeavor centered on data processing technology, aiming to understand and optimize learning, into an interdisciplinary endeavor that broadly captures learning itself through data, providing theoretically meaningful insights through the interpretation and communication of data.

2.1.2 Limitations of LA’s Data-Drivenness

However, it is also clear from the definition that LA is inherently data-driven in nature, despite its evolution. While data-based representations certainly play a crucial role in capturing learners and their contexts, there are also things that cannot be captured and are consequently omitted. The prime example of this is what learners have learned through their learning actions: that is, the learner’s “knowledge.” In practice, the “data about learners and their contexts” handled by LA primarily include observable behaviors and environmental factors, resulting in a situation where the internal cognitive states of learners have rarely been directly addressed in LA. Some also have warned that LAs may consider what is known from the data to be all that is characteristics of learning (Tsai & Gasevic, 2017; Wilson et al., 2017). Therefore, it must be said that LA has a shortcoming in representing the learner’s state of knowledge, since it has not researched the material’s content. This is a major problem, since LA advocates that its goal is to understand “learning and the environment in which it occurs.”

2.2 Knowledge Modeling

Early knowledge models represented knowledge as semantic networks: concepts are nodes and relations are labeled links; inference involves traversing paths and spreading activation (e.g., Collins and Quillian (1969)’s hierarchical model; Collins and Loftus (1975)’s spreading activation). These ideas established that not only frequency but meaningful structure matters for access and reasoning, a premise that underpins today’s graph-based learning analytics.

Knowledge graphs (KGs) generalize semantic networks with explicit schemas or ontologies, typed entities and relations, and often formal semantics (e.g., Resource Description Framework (RDF), Web Ontology Language (OWL)). KGs are now a unifying substrate for integrating heterogeneous educational data, supporting search, recommendation, and

reasoning at scale. Recent surveys characterize KGs as “graphs of data that accumulate and convey knowledge of the real world” (Peng et al., 2023), and review their evolution (from static to temporal/dynamic/event KGs), construction, and reasoning techniques: developments directly relevant to capturing curricular change and learner trajectories.

Within education, systematic reviews report a rapid rise of KG applications (Qu et al., 2024): concept/knowledge-point modeling, curriculum mapping, semantic search/Q&A over learning objects, personalized learning support and program management. These studies note that most educational KGs to date are course- or institution- specific and emphasize engineering over generalizable learner-knowledge semantics.

2.3 Learner Knowledge Representation and LA

Even before the advent of LA, various attempts have been made to objectively represent learners’ knowledge states in learning and educational research. Knowledge space theory (Falmagne & Doignon, 2010) models the learner’s knowledge state based on the relationship between problems and solutions by representing the structuring of knowledge and learning progression. Various knowledge tracing methods, such as Bayesian Knowledge Tracing (BKT) (Corbett & Anderson, 1994) and Deep Knowledge Tracing (DKT) (Piech et al., 2015), have been used to estimate the degree of knowledge acquisition by calculating the probability that a learner will answer each question correctly. Item Response Theory (IRT) (Hambleton & Swaminathan, 2013) measures subject characteristics and the difficulty and discrimination of assessment items based on responses to a series of assessment items. However, all the above theories and methods are based on learner responses to specific problem items and consequently cannot estimate knowledge states based on learners’ daily learning behavior.

Another direct example of a means of representing learner knowledge is the Learner Model (Dillenbourg & Self, 1992), which uses a computer to estimate and represent learner knowledge within the framework of an Intelligent Tutoring System (Nwana, 1990). Open Learner Model (Bull, 2020) has the effect of increasing the accuracy of the model itself and enhancing learners’ metacognitive abilities and self-regulated learning skills (Zimmerman, 2002) by disclosing the LM to learners (Hooshyar et al., 2020). However, inference of knowledge state using OLM is still based on specific actions of the learner, and there have been few cases in which inferences were made based on daily learning activities.

In response to this situation, it has been desired to give birth to a research field that integrates OLM and LA (Bodily et al., 2018).

The limitations of existing research can also be cited in terms of how knowledge components are handled. These above approaches tend to obscure specific knowledge-related structures and learning behaviors. Although these are models that assess knowledge levels based on actions, they do not explicitly address what learners have done (Zanellati et al., 2024). As a result, these approaches suffer from model black-boxing (Piech et al., 2015). Moreover, many of these aim to estimate understanding and rarely explicitly address the relationship between knowledge and actions. Previous research argues that explicitly handling knowledge elements and mapping them to learning behaviors is crucial (Abdelrahman et al., 2023), as it enables estimating what was learned from specific learning behaviors and analysis, visualization, and intervention design. Furthermore, the structuring of knowledge domains and their connection to LA have been emphasized, with some arguing that adaptive support based on knowledge structures lies at the core of innovative learning. These calls align with early visions proposed by Siemens (2013) and Hwang (2014), for example.

In this context, the concept of KALA was introduced by Chen (2019). KALA is a new paradigm for LA centered on knowledge. Aiming to transition from a data-driven approach to a knowledge-driven one, it uses data-driven methodologies to go beyond mere behavior tracking and instead focuses on LA to handle learners’ knowledge. Traditional LA has been considered to lack the knowledge awareness, an explicit consideration of learners’ encounters with structured domain knowledge. As a pioneering study that spiritually inherits KALA’s philosophy, Ain et al. (2024) have attempted to link learning behaviors to knowledge elements using knowledge graphs and interpret learning logs semantically.

2.4 Implications for OKLM

Across traditions, two design tensions recur:

1. **Closed vs. Open Worlds.** Knowledge space and knowledge tracing approaches excel when the domain is fixed. Similarly, KGs excel at open-ended, cross-resource integration but historically under-specify the learner state. A KALA framework should compose these: use KGs to model domain structure and resource semantics;

use probabilistic or information-theoretic learner models to estimate state over KG subgraphs.

2. **Explainability vs. Predictive Power.** Symbolic graphs/ontologies offer transparency, while DKT provides accuracy but opaqueness. Recent hybrid methods (rules over KGs (Wu et al., 2023)) demonstrate that combining symbolic constraints with representation learning yields both tractable reasoning and generalization, which is promising for explainable recommendations and diagnostics in OKLM.

Chapter 3

Design of OKLM

3.1 The OKLM Framework Concept

OKLM can be defined as “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, Koike, et al., 2024). It proposes a framework that maps learning actions to knowledge elements extracted automatically from existing instructional materials to address the challenges mentioned above. By doing so, OKLM enables estimating “what was learned” from each observed learning behavior, without relying on handcrafted knowledge models, which can often be tedious to build (Khadir et al., 2021). Its design does not assume perfect knowledge structures, improving feasibility and scalability.

OKLM captures what learners learn through which actions by linking learning logs collected via LA technology with knowledge elements that constitute existing knowledge models. This feature enables the capture of learning behaviors at a granular level due to the use of LA technology, while also eliminating the need to invest significant effort in constructing elaborate knowledge models. Furthermore, it offers the advantage that if a knowledge model exists, it enables the construction of learner models in any field or domain.

3.2 System Requirements

To realize the characteristics of OKLM listed above, OKLM must maintain the following three aspects of openness: knowledge-domain openness, learner openness, and LA system openness. Each of these opennesses is a system requirement for estimating learners’

knowledge states.

3.2.1 Knowledge-Domain Openness

OKLM is based on LA technology. LA retains numerous interactions between learners and various learning materials as learning logs. Therefore, the knowledge model used in OKLM must be adaptable to diverse learning materials. That is, it must be “open” to various knowledge domains. This characteristic enables OKLM’s major feature: modeling a learner’s knowledge state for any knowledge domain, provided the knowledge model exists.

3.2.2 Learner Openness

OKLM aims to be “open” to stakeholders involved in learning, such as learners and educators. LA is designed to enhance users’ understanding and optimization of learning through data visualization and analysis. In the context of learning, the users are learners and educators. Therefore, the ability for users to visualize and analyze models is a crucial feature for understanding and optimizing learning. This is inspired by OLM. In fact, it has been demonstrated that OLM’s “learner openness” yields learning effects, such as improving model accuracy and enhancing learners’ metacognitive abilities by understanding their proficiency levels (Bull, 2020).

3.2.3 LA System Openness

OKLM must also aim to be “open” to various learning support systems using LA. LA integrates and analyzes learning logs obtained from diverse systems. Therefore, it must be able to treat logs from these various LA systems equally. Furthermore, it is entirely possible that these diverse systems handle knowledge domains that differ, and it is conceivable that the learners using the systems will also be diverse. Therefore, by combining these two aspects of openness, LA system openness must be realized for various learning support systems.

3.3 Mathematical Formulation

This section introduces the mathematical formulation of OKLM. Starting from the mathematical representation of knowledge models and learning logs, we propose a method to

express the knowledge learned by learners and their proficiency levels using mathematical formulas. We also introduce the concept of “learning information content” in ITLAK, which will be introduced in Section 5.2.

3.3.1 Formulation by Relative Proficiency

Definition of Knowledge Models

Human knowledge 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 $w : A \rightarrow [0, 1]$ is a mapping to represent the weights of the branches. The weight of a 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 .

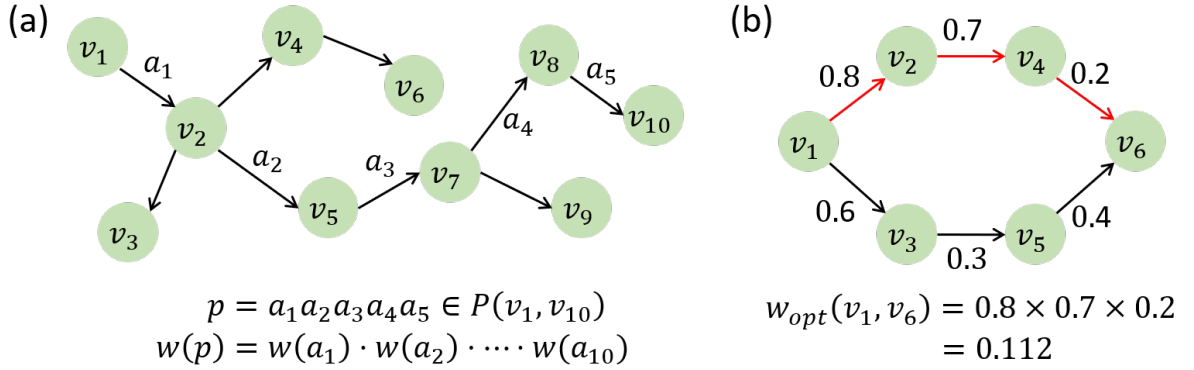


Figure 3.1: (a) A pathway and its weight in the graph structure of a knowledge model. (b) The optimal value of the path weight for the knowledge analogy (Takii, Liang, et al., 2024).

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 $P : V \times V \rightarrow \bigcup_{n \in \mathbf{N}} A^n$. Then, the pathway $P \in P(u, v)$ can be expressed as $p = a_1 a_2 \dots a_n$. The weight of the path p is defined as $w(p) = \prod_{a \in p} w(a)$ (Figure 3.1a). 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 3.1b).

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 .

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 3.2).

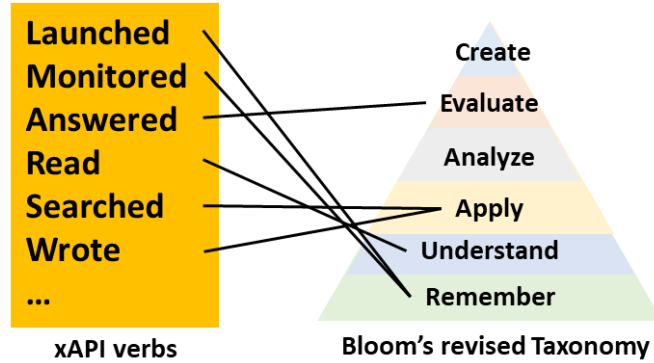


Figure 3.2: Correspondence of xAPI verbs and the stages of Bloom’s revised Taxonomy (Takii, Liang, et al., 2024).

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 τ is denoted as $x_i(s, v, \tau)$, where $i \in I = \{\text{Remember, Understand, Apply, Analyze, Evaluate, Create}\}$.

Definition of 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 τ as follows:

$$rp_i(S, s, v, \tau) = \frac{1}{100} \max \left\{ 0, \min \left\{ \frac{10(x_i(s, v, \tau) - \mu_i(S, v, \tau))}{\sigma_i(S, v, \tau)} + 50, 100 \right\} \right\}. \quad (3.1)$$

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 (3.1) 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%).

3.3.2 Formulation by ITLAK

To apply ITLAK in a concrete learning context, it is necessary to define the information content of learning behaviors mathematically rigorously. This section introduces the core formalization of ITLAK, which quantifies how much information a learner is expected to gain from each learning action based on their prior interactions. We assume that learning actions are logged as discrete events, each of which can be associated with one or more knowledge components. Such an association is made possible through the OKLM, which maps individual learning logs to specific knowledge elements embedded in instructional materials. The knowledge elements in the OKLM are extracted from learning materials in advance and take the form of graph structures, in which nodes are the elements and edges are their semantic connections (see the upper part of Figure 3.3). This foundation allows ITLAK to define a learning action not just as a behavioral trace, but as an observable instance of contact with structured knowledge.

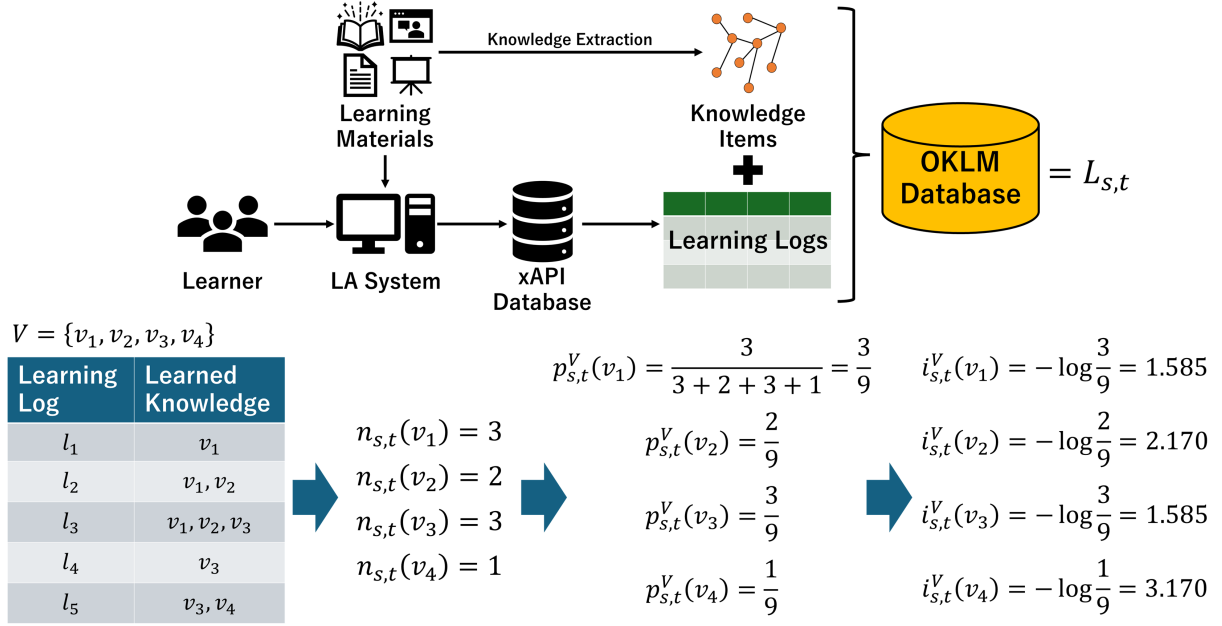


Figure 3.3: An OKLM structure and an example of calculating information content of ITLAK (Takii et al., 2025). Learning logs retrieved from an LA system are linked to knowledge elements, as shown in the table in this figure. The fewer times an element has been encountered, the higher its information content. In this study, we calculate using base 2. Although α is set as 0 in these examples, we guarantee $\alpha > 0$ in an actual setting.

Here, let us assume a learning log as the smallest unit of learning behavior, represented as l . The l can be uniquely determined by the learner who left the learning log and the time it was left. Specifically, in case the learning log l is in the form of Experience API (xAPI) Advanced Distributed Learning Initiative (ADL), n.d., l can be expressed as a tuple of Actor, Verb, Object, Context, Result, and Timestamp, capturing the learner’s interaction with content. We also define $L_{s,t}$ as a set of all learning logs a learner s leaves by time t . To connect learning logs with knowledge elements, $n_{s,t}(v)$ is defined as the number of learning logs in $L_{s,t}$ that are associated with knowledge element v :

$$n_{s,t}(v) = |\{l \in L_{s,t} : v \in k(l)\}|, \quad (3.2)$$

where k is a correspondence that maps a learning log to a set of knowledge elements.

To define the information content of a learning behavior, it is necessary to consider the uncertainty or the probability of the knowledge elements with which the learning behavior is associated. As a measure of this uncertainty, we employ the distribution of the frequency with which a knowledge element appears in a series of learning logs.

Considering a set V of knowledge elements corresponding to the knowledge domain, this is represented as $p_{s,t}^V(v)$ and defined as follows:

$$p_{s,t}^V(v) = \frac{n_{s,t}(v) + \alpha}{\sum_{v' \in V} (n_{s,t}(v') + \alpha)}, \quad (3.3)$$

where $\alpha > 0$ is a smoothing parameter to address the zero-frequency problem. Based on past learning logs, this probability shows a learner's likelihood of engaging with a particular knowledge element. Using the probability, the information content $i_{s,t}^V(v)$, obtained when the learner s has learned a knowledge element v on time t , can be simply defined by the following formula:

$$i_{s,t}^V(v) = -\log p_{s,t}^V(v). \quad (3.4)$$

The lower part of Figure 3.3 shows some examples of the calculation of information content. Assuming that learning logs are written in xAPI format, each learning log is stored in an xAPI log database through the LA system. The xAPI logs stored in the database that learner s has left by time t is treated as $L_{s,t}$. In this figure, learner s has left five learning logs l_1, l_2, l_3, l_4 , and l_5 by time t , and the set of knowledge elements V is assumed to be $\{v_1, v_2, v_3, v_4\}$. The information content for each knowledge item is calculated as shown in the examples. These examples show that $i_{s,t}^V(v_4)$ has the largest value, which means that knowledge element v_4 is the most novel for learner s at time t .

In addition to the information content associated with each knowledge element, it is also meaningful to consider the overall uncertainty of a learner's knowledge engagement at a given time. To capture this, we define the entropy of the learner's behavior distribution over the knowledge set V as $H_{s,t}^V = -\sum_{v \in V} p_{s,t}^V(v) \log p_{s,t}^V(v)$. This entropy $H_{s,t}^V$ represents the degree of unpredictability in the learner's future interaction with knowledge elements, based on their learning history up to time t . Within the ITLAK framework, entropy provides a macro-level indicator of the learner's exploratory diversity or behavioral novelty in their engagement with the knowledge domain. It complements the micro-level analysis offered by the individual information values $i_{s,t}^V(v)$, offering a broader perspective on the structure of a learner's knowledge interaction pattern.

3.4 OKLM Visualization

Finally, we describe the visualization component that enables intuitive understanding of learner knowledge representation in OKLM. While not an essential function of OKLM,

this feature is considered beneficial for users seeking to grasp the learner’s knowledge state.

This visualization tool allows us to select the student name, instructional material name, and indicator name to display OKLM data. We can filter the data within OKLM based on specified conditions to customize the visualization. Selecting a student name enables us to display either an individual student’s OKLM or an overlay of all students’ OKLM. Selecting an instructional material name allows us to extract and display only the knowledge elements contained within that specific material. Furthermore, indicator names represent the types of learning behaviors used as the basis for OKLM’s proficiency calculations. In Figure 3.4, showing the visualization sample of OKLM, “Highlighted with a marker” indicates that proficiency was determined based on whether each English word had been highlighted with a marker. Additionally, moving the slider at the top of the figure left or right allows visualization of OKLM’s time-series changes, specifically the temporal evolution of learner proficiency.

The Figure 3.4 visualizes the learning status of all English words for every learner by registering English vocabulary as a knowledge model. Within this visualization, circular nodes represent knowledge elements (in this case, English words), while the branches connecting nodes represent semantic relationships between words. Green words indicate those judged by learning log analysis to be well-mastered by the learner, while yellow words indicate those the learner does not understand well. Larger green areas indicate higher mastery, while larger yellow areas indicate a greater degree of unfamiliarity, meaning lower mastery.

Chapter 4

System Implementation

4.1 LEAF System

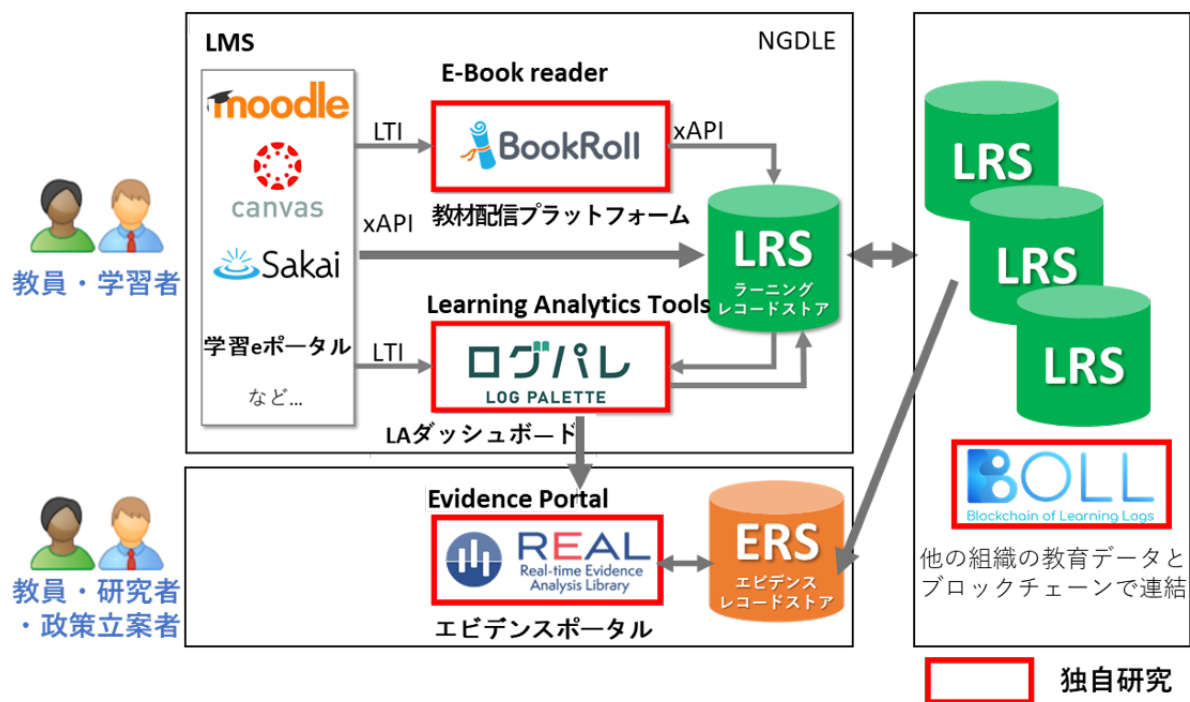


Figure 4.1: Structure of the LEAF system

The OKLM framework requires an infrastructure that can flexibly integrate diverse learning data, support interoperability across systems, and provide a stable environment for experimental as well as production use. To meet these requirements, this research employed the LEAF (Figure 4.1) (Ogata, Majumdar, et al., 2024).

LEAF is an LMS-independent platform designed to connect various educational applications with a centralized LRS and analytics environment. By separating the research

and production layers, it allows for iterative development and testing of new LA methods while ensuring scalability and robustness in real classrooms. Within this thesis, LEAF functioned as the backbone that collected and stored learning logs, managed user authentication, and enabled cross-system data sharing.

Generally, in LA systems, to ensure interoperability, learning records obtained from one learning tool must be correctly interpreted by other learning tools. To meet this requirement, various organizations have established standards to promote Learning Tools Interoperability (LTI). For example, xAPI standard has been proposed for making a statement about learner’s activities withing an LA tool. The xAPI standard uses a structure that has *actor*, *verb*, and *object* to specify how learning statements should be reported so that the semantic and syntactic meaning of learning records are correctly conveyed. Figure 4.2 shows an example of an xAPI statement. An *actor* represents a learner or user who has performed a learning action within the LA system, and this attribute information is recorded in the statements. *Verb* is the type of learning action performed by the *Actor*, and *Object* is the target of that action. With this understanding, we can read the sample xAPI shown in Figure 4.2 as “an Agent *admin* launched a book, whose name is ‘What is Moodle?’ ”

The LEAF system also handles and collects learning logs compliant with the xAPI standard, and therefore the practical advantage of LEAF lies in its interoperability. OKLM was not implemented as a closed system but was instead embedded within LEAF, making it accessible to external applications such as e-book readers, grammar learning systems, and recommendation engines. This openness reflects one of the central design principles of OKLM itself—system openness—and enabled the empirical studies reported in this thesis. Moreover, by situating OKLM within LEAF, the research ensured that the proposed models could be tested under realistic educational conditions rather than in isolated experimental settings.

In summary, LEAF provided the technical foundation that made it possible to validate OKLM across multiple domains and systems. It operationalized the theoretical vision of openness by allowing knowledge models, learner models, and external LA tools to interconnect within a single ecosystem. Without this infrastructure, the demonstration of OKLM’s scalability and interoperability would not have been feasible.

```

{
  "version" : "1.0.0",
  "actor" : {
    "objectType" : "Agent",
    "account" : {
      "name" : "admin",
      "homePage" : "https://www.example.org/bookroll"
    }
  },
  "verb" : {
    "id" : "https://w3id.org/xapi/adb/verbs/launched"
  },
  "object" : {
    "objectType" : "Activity",
    "id" : "https://www.example.org/bookroll/book/view?contents=1234",
    "definition" : {
      "name" : {
        "en" : "What is Moodle?"
      },
      "description" : {
        "en" : "foobarbaz"
      },
      "extensions" : {
        ...
      }
    }
  }
}

```

Figure 4.2: An example xAPI statement

4.2 System Components

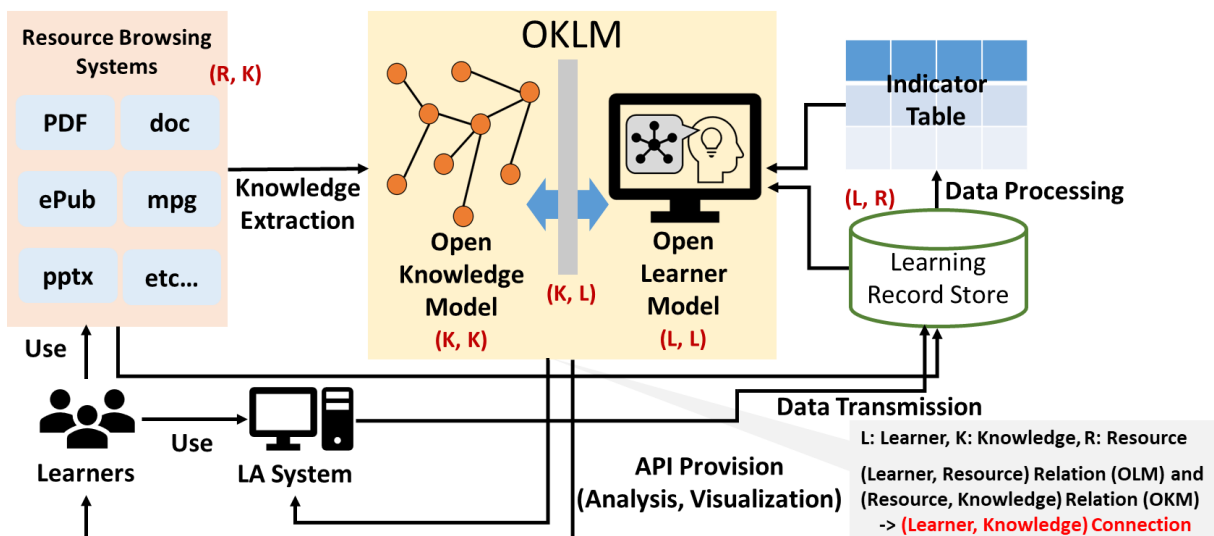


Figure 4.3: Framework of OKLM

As shown in Figure 4.3, OKLM includes the following two models as components. One is a knowledge model, a graph structure representing the knowledge items in the electronic learning materials. The other is a learner model, which is generated from the usage logs of the LA and the learning material browsing systems, and represents the learner’s knowledge state. The graph structure that makes up the knowledge model represents many knowledge items and the syntactic or semantic relationships that exist among them. Each node of the knowledge model contains information on learning activity log data collected from the LA and other learning material viewing systems. Figure 4.4 shows a sample of internal data in OKLM. a knowledge model and a learner model are implemented using Neo4j, a graph database well-suited for graph structures. Furthermore, since OKLM depends only on the knowledge model created from the learning material, the learning material from which the knowledge model is built can be in a variety of formats, such as PDF, PowerPoint, or video.

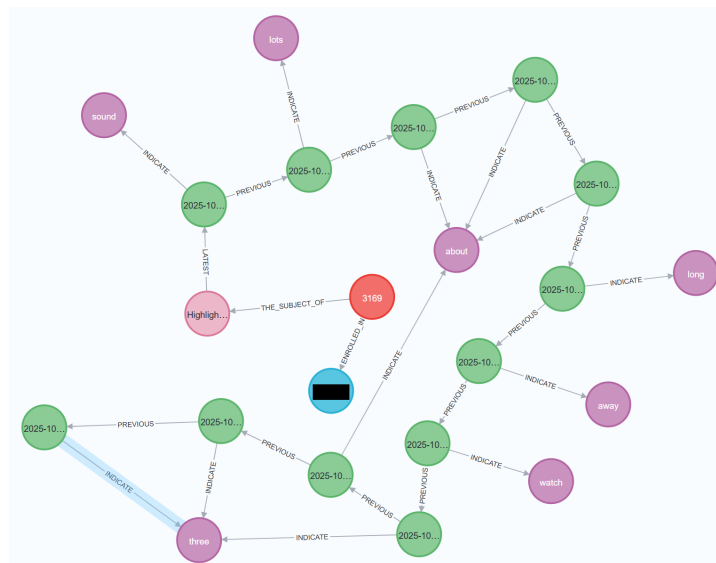


Figure 4.4: Sample of OKLM internal data. Visualization utilizes Neo4j function.

OKLM connects learning activities in the LA system to knowledge information in learning materials. First, various knowledge items are extracted from the learning materials, and a knowledge model is constructed based on their syntactic relationships. The knowledge items can be extracted using various methods developed by previous studies, such as Flanagan, Majumdar, et al. (2019). Second, the learning logs collected from the LA system generate a correspondence between the learner’s learning activities and the

materials that target them. It is possible to generate relations between learning activities and knowledge items using these two types of relations. In other words, estimating which knowledge a particular learning activity learns is possible.

Compared to traditional learner modeling technologies, such as overlay and buggy approaches (Nwana, 1990), OKLM builds knowledge models by extracting and reconstructing knowledge from learning materials using existing methods, whereas they need the process of knowledge modeling that comprehensively describes the knowledge, which often requires expertise and experience (Wang et al., 2017). Although the accuracy of the knowledge models in OKLM may be inferior to that of existing LMs, OKLM compensates for its shortcomings with the sheer volumes of learning activity data and guarantees the accuracy of the LM. This data-driven approach also guarantees the knowledge-domain and LA system openness of the OKLM, as mentioned above.

Chapter 5

Experimental Studies

5.1 Study 1: Knowledge Modeling & Simple Learner Modeling

The following sections describe the details of the OKLM implementation and validation experiments conducted to address the RQs initially presented. First, in this Study 1, we conducted an empirical study to verify the effectiveness of knowledge modeling within OKLM and the management of the created knowledge models. Here, using a knowledge model created for English vocabulary and a simple learner model, we developed the foundational mechanism for OKLM. The developed mechanism was utilized for explainable English extensive reading (ER) material recommendation. We also investigated the relationship between learners' usage of the recommendation and their characteristics as learners.

5.1.1 Research Proposals and Research Questions

The purpose of this study is to create a knowledge model that forms the basis of OKLM and to verify its effectiveness. To achieve this, we created the knowledge model by measuring the difficulty level of each educational resource using the ENglish vocabulary contained within ER materials and their respective difficulty levels.

The recommendation proposed in this study is intended to be a persuasive recommendation that adapts to the learner's preference and encourages spontaneous learning by the learner. However, as indicated above, this requires an accurate estimation of the difficulty level of the material. In addition, for this ER material recommendation to be accepted by more learners and to support learning effectively, it is necessary to understand the

characteristics of the learners who supported it. Therefore, we pose the following two research questions corresponding to the above two defined challenges.

RQ1-1: To what extent does the proposed mechanism for estimating the difficulty level of English materials for ER estimate the difficulty level of the materials?

RQ1-2: What characteristics of learners supported the recommendation of English language materials for ER that could be explained?

Positioning of This Study

Together with the context of EFL learning, this study aims to provide ER material recommendations tailored to learners' preferences regarding the difficulty of the material rather than estimating learners' knowledge states. This is because estimating learners' knowledge states from their learning behaviors is a difficult task. The EFL material recommendation in this study uses a content-based method, which uses information retrieval techniques to automatically infer the difficulty level of the material and the learner's preference for it. This allows the recommendation system to simultaneously model the learner's state and material knowledge. In addition, the recommendation system also aims to ensure transparency within the system by explaining the reasons for recommending the material. It is hoped that this feature of the recommendation system will enable learners to trust the system and consequently increase their motivation to learn.

5.1.2 Recommendation Platform

Recommender Overview

Figure 5.1 shows an overview of the use of the recommendation platform. First, the platform administrator prepares a vocabulary list and teaching materials from the teaching materials store in advance. This word list contains vocabulary difficulty information and is used to evaluate the difficulty level of the teaching materials. Next, learners use the educational materials on the eBook reader system. This system converts the reading logs of the materials into the format of the xAPI, a comprehensive repository of learning/educational records. The reading logs stored in the LRS, along with the difficulty level information for each material, should be designed to be sent to the LRS, a comprehensive repository of learning/instructional records. Next, the recommendation mechanism generates the materials to be recommended to the learner, along with explanations of

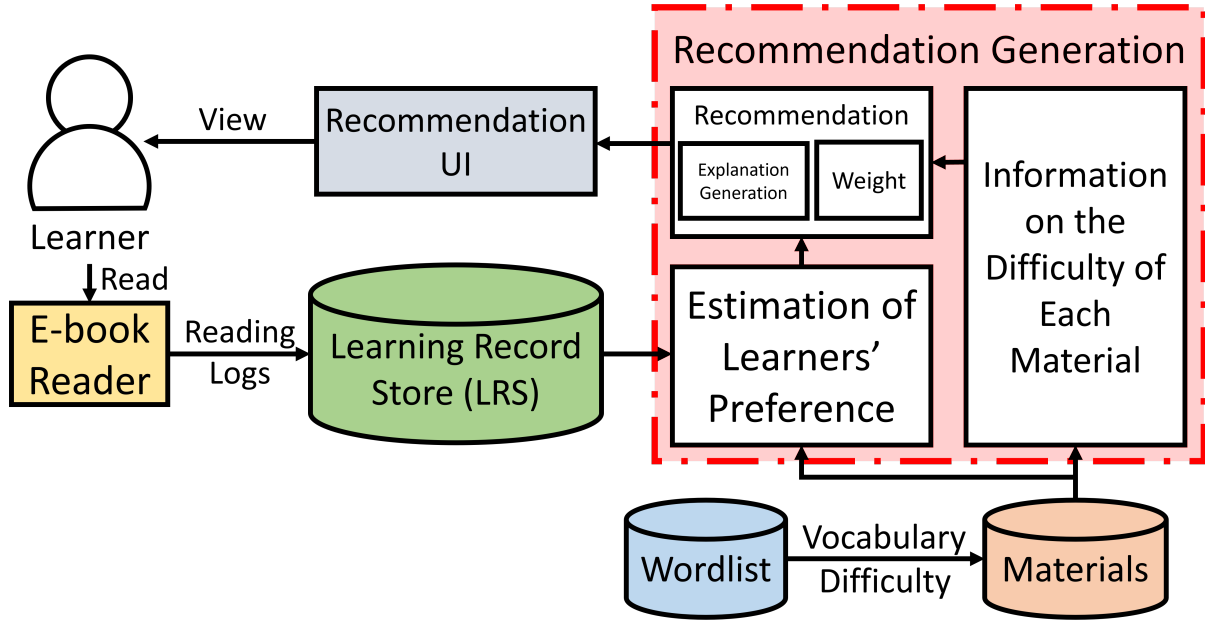


Figure 5.1: An overview of the recommendation platform

the reasons for the recommendations. These explanations include the weight of the recommendation (how much the material should be recommended) and the reasons for the recommendation. Finally, the generated recommendations are presented to the learner through the recommendation interface.

Platform Components

Students can do the ER activity using the BookRoll. With the help of BookRoll, students' reading behaviors are collected, including the reading pages, words, time, and speed. The user interface for the recommendation shown in Figure 5.2 is implemented on the Goal-Oriented Active Learning (GOAL) system (Yang et al., 2024), one of the tool included in the LEAF system. The GOAL system is a platform to support students' development of data-informed self-directed learning (SDL) ability (H. Li et al., 2021; Majumdar et al., 2018). It provides five recommended learning materials at most. Users can jump directly to the BookRoll by clicking the title and can read the recommended material. Each recommended material is followed by the recommendation weight and explanatory sentences explaining why the recommendation was made to the learner. The provided explanation can be shown or hidden when learners click a button next to the explanation.



Figure 5.2: UI of the picture book recommender system implemented in the GOAL system

5.1.3 Recommendation Mechanism

Mechanism Overview

The recommendation mechanism we propose in this study uses model-driven and data-driven approaches. First, as a model-driven approach, this mechanism uses a wordlist with information on the difficulty of vocabulary, which is used to evaluate the difficulty of materials in the book library. Next, as a data-driven approach, reading logs estimate learners' preferences for the difficulty levels of materials with the estimated material difficulty. Then, the system makes material recommendations based on the material difficulty and the learner's preference. This mechanism generates recommendations so that the difficulty of recommended materials will be as close to the learner's preference as possible. Besides, the descriptions which explain why the material has been recommended to the learner are displayed.

Material Difficulty Evaluation

Difficulty of Vocabulary The difficulty level of instructional materials is determined by several factors that make up the materials. In this study, we focus on vocabulary difficulty and assume that vocabulary primarily determines the difficulty of instructional materials. In other words, we assume that teaching materials containing complex vocabulary are generally tricky, and teaching materials containing a few problematic or

many easy vocabulary words are easy to read. However, if the complex vocabulary is not essential for understanding the material, learners can understand the content without knowing the meaning. Therefore, to estimate the difficulty level of the material, we used the vocabulary profile and the ranking function in information retrieval.

This study used the CEFR-J Wordlist Version 1.6 Tono, 2020 as a reference for vocabulary difficulty. This vocabulary list is based on the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2025). It is a corpus created from English textbooks used in China, Taiwan, and Korea, from which common vocabulary used in CEFR-level texts in each country/region is extracted. It was constructed for English language education. The list contains 6868 headwords, each of which has four levels of difficulty: A1, A2, B1, and B2 of the CEFR.

To calculate the difficulty level of the material, we first quantified the difficulty level of each word as A1 to 1, A2 to 2, B1 to 3, B2 to 4, and so on. Since words with multiple meanings have multiple levels for each meaning, these difficulty levels were set as multiple averages. For example, “will” has difficulty level A1 as a modal auxiliary verb and level B2 as a noun meaning mental power. The difficulty is 2.5 in this case, with an average of 1 (A1) and 4 (B2). The exact words are also considered different words if they are spelled differently (e.g., “color” and “colour,” both with difficulty A1). Therefore, the difficulty of the word t , represented as $D(t)$, should be expressed as $D(t) \in [1, 4]$.

Difficulty of materials The difficulty of the documents was calculated using each word introduced above, the TF-IDF score (Rajaraman & Ullman, 2011), and the number of words comprising each document. In our method, all words in each document are extracted and the document is regarded as a bag of words. TF-IDF is a ranking function defined for a tuple of documents and words in information retrieval, which can be applied to bags of words (sets of words in which the elements are allowed to be duplicated) and estimates the relevance of a document to a query. The TF-IDF score refers to the relevance of a query word to a document and indicates the importance of the word in the document. In this study, all words in the “CEFR-J Wordlist Version 1.6” are assumed to be query words. The document score is interpreted as the importance of the word: If the TF-IDF score for the document d and the word t is high, t is an essential word for understanding the contents of d . Besides, we deal with each material as a bag of words consisting of the words in the vocabulary list. Namely, when a set of all words in the vocabulary list is

represented as Σ , the material d is expressed as $d' \cap \Sigma$, where d' is a bag of words of the material corresponding to d .

The difficulty of a material d is calculated by using the following formula:

$$D(d) = f \left(c_1 \sum_{t \in d} D(t) \cdot \text{TFIDF}(d, t) + c_2 \frac{|d|}{\max_{d' \in L} |d'|} \right) (f(x) = \log_{10}(1 + x)), \quad (5.1)$$

where the $\text{TFIDF}(d, t)$ is the TF-IDF score for the tuple of the material d and the word t as a query, L is a set of all the materials, and c_1 and c_2 are constants. This formula can be interpreted as the difficulty of each word and its significance in a particular document contributing to the overall difficulty of the document. Besides, since the number of words composing the material is considered to affect its difficulty (the more words it includes, the more difficult it is), the value obtained by normalizing it is also used. Thus, the difficulty of the document will be computed as a total of products of the difficulty of the word and its significance.

Learner's Difficulty Preference Estimation

Learners generally use English language materials of a difficulty level appropriate to their EFL proficiency. Previous research has confirmed a linear relationship between reading proficiency and the percentage of vocabulary in English language materials (Schmitt et al., 2011). As mentioned earlier, it is challenging to determine proficiency from learners' use of English language materials. However, this fact suggests that the difficulty of the materials used by learners is a good indicator of their EFL proficiency. In this study, learners' use of the learning materials was employed to estimate their preferences for the difficulty level of the materials.

According to the difficulty of the materials mentioned above, we extracted the top five most difficult materials from those used by the learners. We believe the difficulty of these five materials will best reflect the learner's difficulty preference because learners are expected to change the materials they use as their EFL proficiency improves. Then, the average difficulty of these five materials is calculated as the learner's difficulty preference. From now on, the difficulty level preferred by the learner s will be denoted as $P(s)$. Namely by denoting the five most difficult materials as d_1, d_2, d_3, d_4 , and d_5 , $P(s)$ can be calculated as follows:

$$P(s) = \frac{1}{5} \sum_{l=1}^5 D(d_l). \quad (5.2)$$

This feature guarantees that the value of $P(s)$ is close to the maximum value of the learner’s estimated difficulty preference.

Recommendation Generation

Explanation for the Recommendation Our proposed recommendation algorithm is designed to provide feedback to the learner on the recommended material and the reason for recommending it. This allows the learner to be convinced of the recommendation and motivated to learn. This supplementary information includes the learner’s difficulty preference (it implicates his/her proficiency), the recommended material’s weight, and a sentence explaining why the material suits the learner.

Table 5.1: Sentences that explain why the materials were recommended. They depend on the difference between the difficulty of the material and the learner’s difficulty preference.

$D(d) - P(s)$	Explanatory Sentence
less than -0.3 (very easy)	“This book is easy, but you can learn basic vocabulary with fun from this.”
-0.3 to -0.1 (easy)	“This book is a little easy, but you can learn important vocabulary with this book.”
-0.1 to 0.1 (average)	“This book is perfect for your English skills!”
0.1 to 0.3 (difficult)	“This book is a little difficult but worth trying!”
more than 0.3 (very difficult)	“This book is challenging. Let’s give it a try!”

We prepared five types of sentences explaining why the materials were recommended to the learner. According to the theory of proximal learning, when learners decide whether to study, they depend on their belief whether they already know the items. They will choose not to study if they believe they know the item already, and vice versa (Metcalf & Kornell, 2005). Thus, the explanatory sentences should lead the learners to select the recommended materials. The types of sentences depend on the difference between the difficulty of the recommended material and the learner’s preferred English difficulty level, i.e., the value of $D(d) - P(s)$. This takes a value close to 0 when the difficulty of the recommended material d (i.e., $D(d)$) is close to the learner s ’s difficulty preference. When the value is smaller than 0, the recommended material is more accessible than the learner’s level, and

when it is larger than 0, the material is more complicated. The recommender provides different sentences according to the learner’s difficulty with recommended materials, as shown in table 5.1.

These messages were written to motivate learners to use the recommended materials. When the difficulty of the materials is close to the learner’s estimated preference, the explanation states that this material perfectly fits the learner’s English level. Besides, even if the recommended material is too easy or difficult, the recommender explains they are worth using.

Recommendation Weight The recommendation of materials by difficulty level refers to the recommendation of materials whose difficulty level is closest to the learner’s estimated English level or the preference for the difficulty level of the material. In other words, recommended materials should not be too easy or too difficult because materials that are too easy or too difficult will negatively affect learning effectiveness. In this study, we introduce a recommendation weight for material d (denoted as $R(s, d)$), i.e., a value that indicates how much we recommend material d to learner s . We therefore define this value as follows:

$$R(s, d) = \log_{10} \left| \frac{1}{D(d) - P(s)} \right|. \quad (5.3)$$

This formula takes a significant value if the difficulty level of the material d (i.e., $D(d)$) is close to the English preference level of the learner s (i.e., $P(s)$), and a small value otherwise. If the recommendation weight $R(s, d)$ takes a significant value, it will recommend material d more strongly to learner s . This guarantees that the learner s receives the recommendation of the material d , whose difficulty is close to s ’s English skill levels. In the implementation, the value of $D(d) - P(s)$ is adjusted so that it is not zero. The recommendation weights provided are linearly normalized so that the minimum value is 0.0 and the maximum value is 100.0 for ease of understanding by the user. The learner’s preference and the difficulty level of the material are similarly normalized.

5.1.4 Sub-Experiment for Automated Material Difficulty Estimation

Method Overview

The difficulty estimation of English materials presented in this study is essential for presenting how strongly recommended the recommended materials are for each learner. To show that this difficulty estimation mechanism can make accurate estimates, we conducted a simple experiment to measure the difficulty level of English materials. In this experiment, we used English picture books with difficulty levels ranging from A1 to B2/C1 on the CEFR level. Since these difficulty levels were displayed in the picture books in advance, we were able to check whether the formula presented above ($D(d)$) worked by comparing these difficulty level indicators with the difficulty level values estimated by the formula. Therefore, we calculated the difficulty level of the material using the formula and summarized it using the difficulty level labels. The constants c_1 and c_2 were set to 0.1 and 1, respectively.

Result

Table 5.2: Results of an evaluation of an equation that computes the difficulty of books

Difficulty	# of books	Average of Estimated Difficulty
A1	193	0.857
A1/A2	10	1.196
A2	70	1.210
A2/B1	92	1.210
B1	10	1.743
B1/B2	10	1.807
B2	10	1.908
B2/C1	10	1.995

Table 5.2 shows the average number of books and estimated difficulty for each difficulty label. The results show that as the difficulty level of the label increases, the estimated difficulty level also increases. To confirm that the rank order of the estimated difficulty

level coincided with the rank order of the label difficulty level, Kendall's rank correlation coefficient between the two was calculated, with a statistically significant value of 0.9820 ($p < 0.001$). This means that the order of difficulty labeled in the books and calculated in the formula has a strong positive correlation. Therefore, it is concluded that the formulas presented in this study can correctly estimate the difficulty level of the materials based on these characteristics.

Focusing on the estimated difficulty level, it can be read that the increase in the calculated difficulty level with increasing difficulty level is significant for the materials with labeled difficulty levels from A1 to B1, but the increase in the estimated difficulty level with increasing difficulty level is slight for the materials labeled more difficult than B1 compared to the previous. This indicates that the increase in the difficulty level of the overall teaching material is smaller than that of the previous one. This may indicate that the effect of vocabulary difficulty on the overall difficulty of the materials is significant for the materials with low difficulty but decreases as the difficulty level increases.

5.1.5 Setting

Participants and Experimental Settings

This experiment was conducted in a Japanese junior high school with an online learning environment through Moodle (Moodle.org, 2017). In this school, all students had tablet PCs and access to the Internet at home, which allowed us to track and analyze students' learning logs in real time. The experiment involved 120 first-year students and 240 second-year students. The explainable recommendation system was made available to all participants, and all were able to use the recommendations at any time during the experiment. Because junior high school is compulsory education in Japan, we did not conduct a multiple-condition comparison experiment to ensure equal educational opportunities. The first-year middle school students began participating on June 8, 2022, and the 120 second-year students began participating on May 10, 2022. The experiment was completed on July 20, just before summer vacation for first- and second-year students. A summary of the experimental setting is shown in Table 5.3.

Materials

When the experiment was conducted, 534 picture books for the ER program were stored in the library and categorized according to the difficulty level of the CEFR. Participating

Table 5.3: Summary of participants and experimental settings

Grade	# of Participants	Start Date	End Date	# of Days
1st J.H.S.	120	June 8, 2022	July 20, 2022	43
2nd J.H.S.	120	May 10, 2022	July 20, 2022	72

students were free to choose and read at their leisure. Table 5.4 shows the books in the library classified by CEFR level: The most common level was A1 or A1+ (for beginners, 205 books), followed by A2/B1 (for elementary and intermediate students, 126 books). Five books in the N/A level had no label regarding their level.

Table 5.4: The number of picture books classified by CEFR levels

Difficulty	# of books
pre-A1	27
A1 or A1+	205
A1/A2	10
A2 or A2+	75
A2/B1	126
B1 or B1+	37
B1/B2	207
B2	12
B2/C1	10
N/A	5
Total	534

Evaluation

The data obtained in this experiment were evaluated from four perspectives: (1) acceptance of recommendations according to learning behavior, (2) acceptance of recommendations according to learner perceptions, (3) impact of recommendations on learner learning activities, and (4) characteristics of recommendations according to learner perceptions.

The evaluation phase used usage logs obtained from the BookRoll and GOAL systems,

and a post-poll was conducted after the end of the ER program period. The usage logs were used to evaluate learning behaviors and learning skills. Three indicators used in the experiment were used to evaluate learning behavior: number of pages read by the learners, reading time, and number of words. Learning ability was assessed by the learners' reading speed (words per minute). The results of the post-poll were used to assess learners' perceptions of the recommendations.

Post-Poll

After the ER program was completed, a post-event survey on the recommender system was conducted. The poll was based on the Technology Acceptance Model (TAM) (Park et al., 2012) and aimed to investigate participants' perceptions of the system's usefulness, ease of use, and attitudes. TAM is an indicator commonly used to measure to what extent the system is accepted by users in the context of learning/educational support by information technologies (Granić & Marangunić, 2019). Each perception included three question items, for a total of nine items. All question items were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A summary of the poll questions and their classification in the TAM is presented in Table 5.5.

5.1.6 Result

System Usage and Recommendation Acceptance

First, following the method of Takami et al. (2023), we define the terms “access” and “click” as used in this study. We define “access” as an action in which a learner accesses the GOAL system's educational material recommendation page, and the system displays the five recommended educational materials. We also define “click” as an action in which a learner clicks on the title of a recommended teaching material and jumps to the book roll page. For example, if a learner accesses the GOAL system's recommendation page and clicks on three titles of recommended educational materials, these actions are counted as “one access and three clicks.”

The number of accesses and clicks to the recommended picture book titles were counted to determine the frequency of use of the recommendation system. Figure 5.3 shows the number of accesses and clicks per day on the recommendation page, separately for first graders (left) and second graders (right). Total Clicks in the figure show the number of times an ER book was opened in BookRoll and whether a recommendation was used. Of

Table 5.5: List of questions in the post-poll and their class in TAM. Each question item was rated on a five-point Likert scale.

TAM in e-learning		Question Item
Perceived usefulness	PU1.	The picture book recommendation has improved my learning performance.
	PU2.	The picture book recommendation has improved my learning efficiency.
	PU3.	The picture book recommendation helps me achieve a high level of learning.
Perceived ease of use	PEU1.	Using the picture book recommendation would make it easy to choose picture books to read.
	PEU2.	It is easy to learn how to use the picture book recommendation.
	PEU3.	It is easy to use the picture book recommendation.
Attitude	AT1.	It is good to read books using the picture book recommendation.
	AT2.	I would like to do learning activities with the picture book recommendation again.
	AT3.	I would like to use the picture book recommendation proactively.

the 240 participants, 100 students accessed the recommendation page at least once, and 46 students clicked on the title of the recommended material at least once. Figure 5.3 shows that except for a big spike in mid-June, the frequency of use of the recommendations was not very high. This result may reflect that all participants were introduced to the recommendation function by their teachers and tried using it in mid-June. Furthermore, first-year students used the ER materials without recommendations more frequently during the entire duration of the experiment. Other than that, there also was a small spike just before the end of the experimental period. In contrast, second-year students did not use ER materials without recommendations, except for big spikes in accesses in mid-May and mid-June. The first-year and second-year students had spikes just after the start of the experimental period in common. Table 5.6, which shows the daily averages of accesses and clicks, shows that participants accessed and clicked approximately 4 and 3 times daily, respectively. It also shows that first-year students used the recommendations

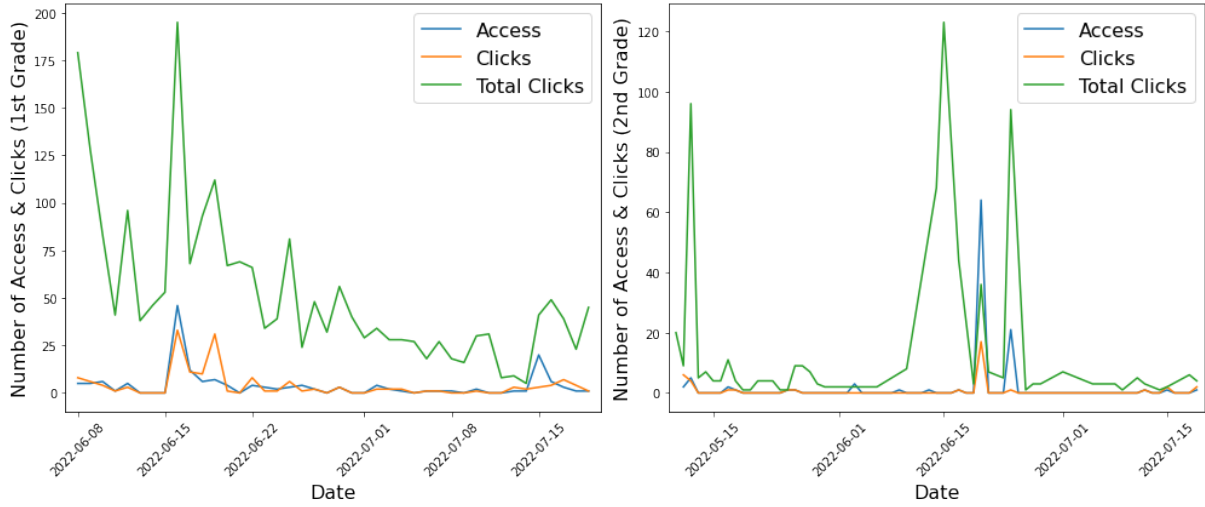


Figure 5.3: Numbers of daily accesses and clicks on the recommendation page

more frequently than second-year students.

Table 5.6: Daily means of the number of accesses and clicks

	# of accesses	# of clicks
1st Grade	3.881	3.881
2nd Grade	1.514	0.543
All	3.843	2.871

Acceptance rates based on the number of accesses and clicks were also calculated to evaluate the degree to which the recommendations were accepted. First, the acceptance rate based on the number of accesses was calculated as the ratio of accesses in which one of the titles was clicked to all accesses. For example, suppose a learner accessed the recommendation page three times: on the first access, he/she clicked on three of the recommendation titles; on the second access, he/she clicked on one of the titles; and on the third access, s/he clicked on none of the titles. Then, the learner acceptance rate based on the number of accesses would be 66.7% ($=2/3$) since none of the titles were clicked twice out of the three accesses.

The acceptance rate based on the number of clicks was then calculated as the ratio of the number of recommended titles clicked on to the total number of recommended titles. In other words, in the above example, the learner clicked on four titles and a total of 15 ($=3 \times 5$) recommended titles, resulting in a learner acceptance rate of 26.7% ($=4/15$)

based on clicks.

The overall pass rate and the pass rate for each candidate were calculated in terms of both accesses and clicks. The results are shown in Table 5.7. This table shows that more than 40% of the accesses led to clicks on recommended books. However, the average acceptance rate for each student was lower than the overall average, indicating that not many students used the recommended books frequently. This implies that many students used the recommended books, but only a few used them continuously and repeatedly. The fact that the number of clicks was lower than the number of accesses also suggests that not many students clicked on more than one title in a single recommendation.

Table 5.7: Acceptance rate of the recommendation. The value was calculated by access to the recommendation or the number of clicks on the title(s) of the recommended books.

	Overall	Student mean
By access	43.49% (=117/269)	26.66%
By click	14.94% (=201/(2695))	10.86%

Acceptance by Learners' Perception

The post-poll results were tabulated for the nine main items shown in Table 5.5. A total of 203 respondents responded to the posterior poll. Cronbach's alpha was calculated to be 0.884 (> 0.8), indicating that the results of the posterior survey were internally consistent and reliable. Table 8 presents the results of the posterior survey by question item.

Table 5.8 shows that the respondents gave high ratings for ease of use (PEU) and attitude (AT) but relatively low ratings for usefulness (PU). Specifically, question AT1, which asked for a general opinion on whether using recommendations to read books was a good idea, had a high score. In contrast, questions PU1 and PU2, which asked how effective and efficient the recommendations were for learning, respectively, had a low score. Although the learners did not feel that the recommendations helped them improve their EFL skills, the results suggest that the recommendations were not complex for them to use. Many learners accepted them favorably and thought using the recommendations was an excellent way to learn English. This suggests that many learners accept recommendations favorably and think that using recommendations to learn English is good. In other words,

Table 5.8: Summary of the post poll results. This shows the number of participants who answered each question as 1 (strongly disagree) to 5 (strongly agree) and the mean values of the answers for each question.

	1	2	3	4	5	Mean
PU1	24	39	102	31	7	2.793
PU2	23	39	104	31	6	2.793
PU3	13	19	78	69	24	3.354
PEU1	17	16	68	73	29	3.399
PEU2	9	18	74	75	27	3.458
PEU3	7	31	80	60	25	3.320
AT1	6	5	58	83	51	3.838
AT2	16	25	83	57	22	3.217
AT3	16	24	70	61	32	3.340

the use of recommendations was effective in motivating EFL learners.

Learners' Characteristics Based on Their Perceptions

Table 5.9: Correlation between learners' activities and poll answers

	Pages	Time	Speed	Words	Accept (access)	Accept (click)
PU1	.120	.105	.094	.118	.008	.009
PU2	.136	.097	.084	.125	.060	.070
PU3	.136	.195**	.001	.155*	.064	.059
PEU1	.180**	.182**	.076	.195**	.032	.042
PEU2	.175	.097	.125	.149*	.128	.128
PEU3	.150*	.112	.149*	.134	.140*	.145*
AT1	.236***	.252***	.127	.263***	.093	.105
AT2	.242***	.282***	.099	.281***	.112	.126
AT3	.254***	.260***	.121	.290***	.120	.122

*: $p < .05$, **: $p < .01$, ***: $p < .005$

To examine learners' attitudes toward and characteristics of recommendations, Spearman's correlations between the learners' learning activities and indicators of their use of recommendations (number of pages read, time spent reading, speed of reading, number of words read, and acceptance rate by number of accesses and clicks) and the results of the posterior survey were calculated. Table 5.9 shows these correlations, which reveal that learners' awareness of the recommendations is significantly and slightly positively correlated with the number of pages, time, and words read, indicators of learning activity. Acceptance, however, showed little correlation with learners' awareness of recommendations, regardless of the number of accesses or clicks. This result suggests that learners who were highly motivated to engage in reading activities had a more positive attitude toward using recommendations than those who were not, which is consistent with the inferences made by the previous results.

Table 5.10: The clustering result using the poll answers (mean (std.)). It also shows the breakdown of answers and learners' activities by cluster.

	Cluster 1	Cluster 2	Cluster 3
PU1	1.694 (0.786)	2.722 (0.639)	3.593 (0.812)
PU2	1.722 (0.779)	2.769 (0.635)	3.492 (0.858)
PU3	2.139 (0.961)	3.370 (0.705)	4.068 (0.848)
PEU1	1.806 (0.889)	3.519 (0.730)	4.153 (0.715)
PEU2	2.861 (1.334)	3.278 (0.747)	4.153 (0.665)
PEU3	2.861 (1.334)	3.065 (0.714)	4.068 (0.763)
AT1	2.889 (1.190)	3.731 (0.650)	4.576 (0.532)
AT2	1.778 (0.832)	3.148 (0.609)	4.220 (0.671)
AT3	1.889 (0.979)	3.259 (0.741)	4.373 (0.613)

For a more detailed analysis, the participating students were clustered using their responses to the post-event poll, and the mean of each indicator was calculated. The clustering algorithm used was K-means clustering. Table 5.10 presents the results of this analysis. First, the number of clusters was determined to be three using the elbow method (see Figure 5.4). The clustering results showed that students in Cluster 1 were the most negative about the recommendation, students in Cluster 3 were the most positive, and

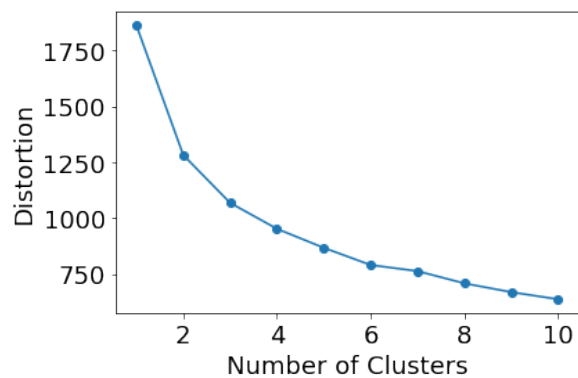


Figure 5.4: Explained variance for the clustering. Based on this, we decided on three clusters.

The Breakdown of Answers and Learner's Activities by Each Cluster

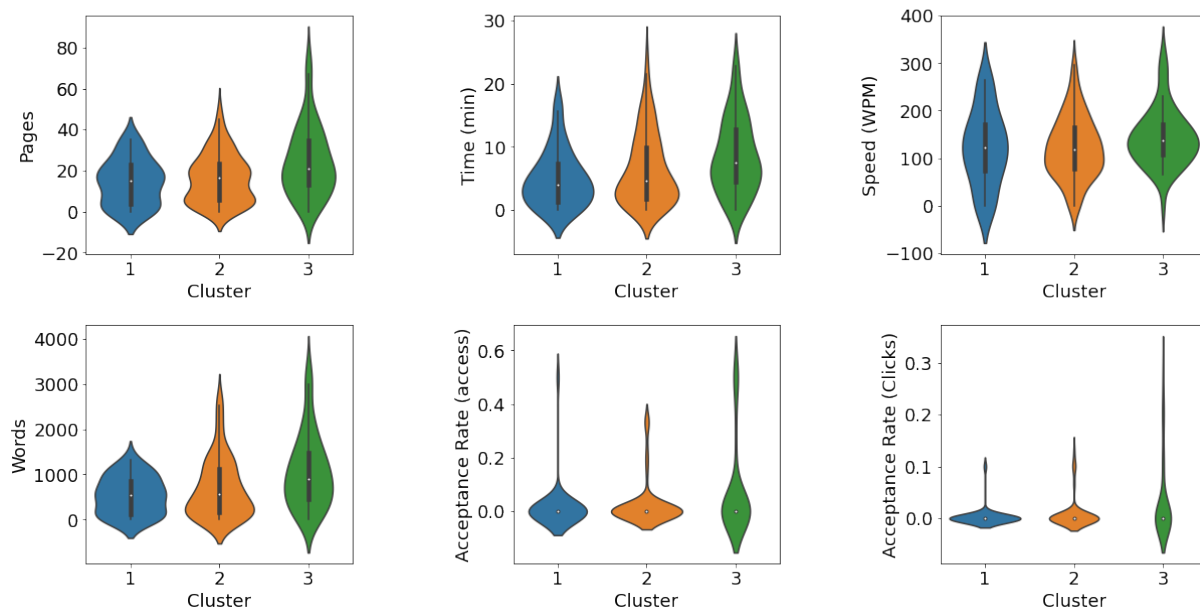


Figure 5.5: The breakdown of answers and learners' activities by each cluster

students in Cluster 2 were in between. Figure 5.5 shows that the values for each indicator are larger for clusters 1, 2, and 3, in that order, except for the pass rate by access. The table shows that students with more positive perceptions are more active in reading activities and read recommended books more often than those who do not. However, the acceptance rate by number of accesses was highest in Cluster 1, slightly lower than in Cluster 3, and lowest in Cluster 2. The fact that students with negative perceptions showed the highest acceptance rate by the number of accesses suggests that the need for recommendations is high but less effective in motivating learners to read. Students in Cluster 3 were probably more engaged in reading prior to the experiment, and the high click rate suggests that they were technologically familiar and found the multiple book recommendations helpful.

5.1.7 Discussion

RQ1-1: To What Extent Does the Proposed Mechanism for Estimating the Difficulty Level of English Materials for ER Estimate the Difficulty Level of the Materials?

The estimation results of the difficulty level of English materials showed that for relatively easy materials, the vocabulary level included tends to affect the overall difficulty significantly. In contrast, the effect of vocabulary is less significant for hard materials. As examples of the components of the difficulty of EFL materials, unfamiliar vocabulary (Kasim & Raisha, 2017; Qrquez & Ab Rashid, 2017) and unfamiliar content and grammatical matters (Satriani, 2018) can be raised. Considering that vocabulary and content, rather than vocabulary itself, has a more substantial influence on the difficulty of the material in complex materials. Therefore, in response to the RQ, the present difficulty estimation system can accurately measure the difficulty of materials that are easier than those that are more difficult to some extent.

Although the difficulty estimation was conducted for ER EFL materials in this study, the results can be generalized to other contexts if the materials and books are mainly composed of English texts. This difficulty estimation mechanism is entirely dependent on the content of the materials and is not related to the intended purpose of the materials. Therefore, measuring the difficulty of commercially available materials and texts written by EFL teachers, for example, is possible. However, it should be noted that this estimation is based on the CEFR-J, which, as mentioned earlier, is vocabulary-dependent and a

difficulty level for EFL learners. In this study, the difficulty estimation was based on materials with pre-labeled difficulty levels. However, this difficulty estimation mechanism also allows us to recommend EFL materials that do not have pre-labeled difficulty levels.

However, there is room for improvement in the method used to generate explanations for recommendations based on the difficulty level of the material. In our work, we arbitrarily set the ranges that define which explanatory sentence is provided to learners. Therefore, it is difficult to evaluate recommendations quantitatively without a learner's subjective evaluation of the recommendation. We require a more detailed investigation to verify whether these values can really serve as a criterion to define a range of difficulties.

RQ1-2: What Characteristics of Learners Supported the Recommendation of English Language Materials for ER That Could Be Explained?

The results of the field experiment in the environment for education showed that few students used the recommendations repeatedly or continuously. On the other hand, students who used the recommendations were more active in their learning activities, suggesting a positive correlation between students' learning activities and the use of the recommendations. It also suggests that many students considered using recommendations a good thing and had a favorable impression, especially regarding ease of use. This idea is supported by the fact that the frequency of use increased only when the teacher introduced the recommendations to the students. Furthermore, although small, students who were highly motivated to read used the recommendations repeatedly, clicking on multiple titles from a single recommendation. Thus, although the recommendations did not motivate much learning, they were attractive, especially to highly motivated students to read.

Comparing the use of the recommendation system by participating first-year and second-year students, more first-year students accessed the recommendations than second-year students. This can be attributed to the ability of the system and the learning with the system itself to arouse the learners' curiosity; the e-learning system itself has the power to arouse the curiosity of the learners themselves (Sarac et al., 2022). The learners' curiosity can explain the small spikes just after the start of the experimental period; they may be curious about the newly introduced function and tried using it. Besides, since first-year junior high school students who have just entered the school are unfamiliar with e-learning systems, it can be assumed that they were more interested in learning with the system and engaged in more learning activities. However, the effect of the system did not

last long and did not arouse much curiosity in the second graders.

5.2 Study 2: Learner Modeling

This study constructs a learner modeling mechanism using OKLM and verifies its effectiveness. Furthermore, to verify LA system openness, the creation of the knowledge model and learner model is conducted in a different educational context from Study 1. The learner model was created using ITLAK, as described in 3.3.2. This study also verifies the ability of ITLAK itself to capture learner characteristics.

5.2.1 Research Proposals and Research Questions

In our research, calculating information content in ITLAK is an attempt to gain insight into the value of knowledge that cannot be computed using behavior-based LA alone. According to information theory (Shannon, 1948), events that occur frequently are interpreted as having low information content, while those that occur infrequently are construed as having high information content. When applied to learning, this suggests that new knowledge is received with surprise or impact, while revisiting knowledge that has already been learned multiple times offers little additional value. Information content in ITLAK theoretically reflects the value of knowledge that differs for each learner. Thus, ITLAK provides a framework for measuring the value of knowledge that cannot be captured by simply counting the times it has been encountered.

In this study, we applied ITLAK to the intensive reading (IR) context. We investigated the relationship between the information content learners obtained for each grammar item and their self-assessment of their translations. Specifically, we prepared an IR environment focused on grammar in EFL and conducted an experimental case study in a real-world setting, targeting junior high school students in Japan. In this environment, open-ended English translation tasks were performed, and detailed learning logs were obtained each time learning activities occurred. Using these logs, we calculated the information content acquired for each grammar item and examined its relationship with learners' self-assessment scores and its explanatory power. To evaluate the effectiveness of this measure, we also compared it against a baseline metric: the simple frequency of contact with each grammar item.

This study serves as a case study of ITLAK and demonstrates the potential of KALA to capture aspects of learning that remain invisible in traditional behavior-focused analytics. As educational environments increasingly integrate AI-based recommendation

systems and knowledge-aware platforms, the demand for explainable, interpretable, and theoretically grounded learning metrics has grown (Gunasekara & Saarela, 2025; Ogata, Flanagan, et al., 2024). ITLAK responds to this demand by offering a principled way to quantify the epistemic value of learning behavior.

This study investigates whether the information content calculated by the ITLAK framework can meaningfully contribute to interpreting learners’ self-assessment scores regarding English grammar proficiency. In particular, we compare the explanatory power of ITLAK’s calculated information content with a more straightforward metric: the number of contacts (i.e., how many times a learner has encountered a given knowledge element). We formulated a central research question (Main-RQ) to address this question comprehensively.

RQ2: To what extent does the information content calculated by ITLAK based on learning logs contribute to interpreting learners’ self-assessment scores regarding English grammar proficiency?

The necessity of considering multiple levels of analysis arises from the multifaceted nature of learning behaviors and knowledge acquisition. Corresponding to these different levels, we formulated three sub-research questions (RQ2-1 to RQ2-3).

RQ2-1: At the level of individual learning logs, to what extent can the information content calculated by ITLAK explain learners’ self-assessment scores?

RQ2-2: At the level of cumulative learning progress, to what extent can the cumulative information content calculated by ITLAK explain learners’ self-assessment scores? In particular, how does its explanatory power differ from simple contact frequency?

RQ2-3: For each knowledge element, to what extent can the cumulative information content calculated by ITLAK explain learners’ self-assessment scores regarding that specific element? In particular, how does its explanatory power differ from simple contact frequency?

First, at the level of individual learning logs (micro analysis), we seek to capture the immediate informativeness of each learning action and its direct relationship with learners’ self-assessment scores (RQ2-1). Second, at the level of cumulative learning progress (macro analysis), we aim to examine how the accumulation of information across multiple learning actions correlates with the broader trend of learners’ overall proficiency development (RQ2-2). Finally, at the level of individual knowledge elements ((s, v) analysis), we investigate whether the accumulation of information specific to each grammar item can

explain item-specific proficiency scores (RQ2-3). By analyzing these three complementary levels, we aim to rigorously evaluate the explanatory power of ITLAK from multiple perspectives and validate its theoretical significance beyond that of simple contact-based metrics.

5.2.2 Experimental Setting

Experimental Period and Participants

This study was conducted with third-grade students at a junior high school in Japan over five weeks from February 6 to March 13, 2025. The subjects were students who participated in the IR system as part of their English classes, and their participation was completely voluntary. All participating students signed a document agreeing to the collection and analysis of data, and all data obtained were anonymized. Given its limited scale, the experiment is treated as an exploratory case study intended to demonstrate feasibility rather than to establish generalizable effects.

IR Environment

Learning took place through the IR system provided on the Moodle platform. This IR system is designed to enable students to study on their own, without the need for teachers during the learning process. This system was permanently installed in Moodle’s English course and made available to students who wished to use it anytime. Students could access the system and work on the provided reading tasks anytime.

Learning Materials and Grammar Items

The English sentences used as reading materials were generated using OpenAI’s ChatGPT (OpenAI, 2025). When generated, the sentences were designed by experimenters to be around 50 words in length, so that native Japanese speakers would have some difficulty understanding them. Initially, 30 questions were provided, and later 50 more were added (with an additional date of February 10, 2025). All English sentences are tagged with multiple English grammar items based on a CEFR-J Grammar Profile. This allows each study log to be considered a contact with a specific grammar item to calculate the information content with ITLAK.

Proficiency Metrics

In the workflow, learners were asked to enter a 7-point self-assessment score for each grammar item after learning. These scores reflect the evaluations of their understanding of each grammar item, based on comparison with the model answers. Although self-assessment is subjective, it has been widely used in metacognitive and language learning research as a proxy for learners’ perceived understanding and has shown moderate validity in capturing perceived proficiency (Brantmeier, 2005; Goto Butler & Lee, 2010). In this study, we treat this self-assessment score as an indicator of the learner’s proficiency in the grammar and analyze its relationship to the information content calculated by ITLAK and the number of contacts.

5.2.3 Analysis Method

To answer the research question of this study, we conducted three types of analysis: micro-level analysis, macro-level analysis, and learner-and-knowledge-based analysis ((s, v) analysis). First, in the micro-level analysis, we analyzed how each learning activity of answering a single English IR question was related to the acquisition of grammar items. This can reflect the immediate effect of a single learning activity. Next, in the macro-level analysis, we analyzed all learning behaviors performed by learners during the entire learning period and the proficiency of grammar items. We also compared this cumulative effect with that of simple contact frequency. Finally, in the (s, v) analysis, we analyzed how the cumulative information content and contact frequency for each grammar item were related to learners’ proficiency scores for that specific item. This can mean the effect of learning particular to each knowledge element. Unlike simple frequency counts, the information content defined in ITLAK captures the epistemic novelty of each learning action, making it a more theoretically grounded indicator of cognitive engagement. Table 5.11 shows a list of notations used in this study. The statistical quantities measured and analyzed in this study are represented using these symbols. Also, α in equation (3.3) was set as 1.

Micro Analysis: Immediate Effect

In the micro analysis, we considered each learning log l , uniquely identifiable by a tuple of a student and time (s, t) , as an independent unit of analysis. For each log, we calculated

Table 5.11: Notation Used in This Study

Symbol	Definition
s	A learner
t	A timestamp (time of a learning log)
v	A knowledge element (e.g., a grammar item)
$l(= l(s, t))$	A single learning log generated by learner s at time t
$L_{s,t}$	The set of learning logs generated by learner s up to time t
$k(l)$	The set of knowledge elements involved in learning log l
$n_{s,t}(v)$	The number of learning logs in $L_{s,t}$ that involve knowledge element v
$p_{s,t}^V(v)$	The probability that a learning log involves knowledge element v
$i_{s,t}^V(v)$	The information content of knowledge element v at time t
$\text{CumInfo}(s, t, v)$	The sum of information content related to knowledge element v accumulated up to time t
$\text{CumCount}(s, t, v)$	The total number of contacts with knowledge element v accumulated up to time t
$\text{GR}(s, t, v)$	The self-assessment score related to grammar proficiency

the mean information content $\bar{i}_v$ over the set of grammar items $k(l)$ involved, defined as

$$\bar{i}_v(l) = \frac{1}{|k(l)|} \sum_{v \in k(l)} i_{s,t}^V(v). \quad (5.4)$$

I also computed the mean self-assessment score over the same set of grammar items, represented as

$$\text{GR}_{s,t} = \frac{1}{|k(l)|} \sum_{v \in k(l)} \text{GR}(s, t, v), \quad (5.5)$$

In plain terms, equation (5.4) calculates how “novel” or “informative” the current learning log is, by averaging the information content over all grammar items involved. Equation (5.5) captures the learner’s overall perceived understanding of those items.

I then examined the relationship between these two variables and the explanatory power of the information content by calculating Spearman’s rank correlation coefficient and using simple linear regression, respectively. This analysis captures the immediate effect of individual learning actions on learners’ perceived proficiency. This analysis focuses on the short-term impact of novelty or surprise in each learning instance.

Macro Analysis: Cumulative Effect

In the macro analysis, we examined how the cumulative learning behaviors of each learner relate to their grammar proficiency. This macro analysis aims to capture the cumulative effect of learning over time. By segmenting learning logs into time-ordered windows, we strive to evaluate the difference between the explanatory power of the accumulation of information content, representing sustained exposure to diverse knowledge, and raw contact frequency regarding grammar proficiency. Let the learning logs of learner s be chronologically ordered as $l_1, l_2, \dots, l_N$, with each log l_i recorded at timestamp t_i and involving a set of grammar items $k(l_i) \subseteq V$. To analyze the temporal progression of learning, we segmented the sequence into non-overlapping windows of fixed size w , yielding a sequence of windows:

$$W_j = \{l_{(j-1)w+1}, \dots, l_{jw}\}, \quad j = 1, 2, \dots, M = \left\lfloor \frac{N}{w} \right\rfloor. \quad (5.6)$$

For each window j , we computed the cumulative information content up to window j :

$$\text{CumInfo}_j = \sum_{i=1}^{jw} \sum_{v \in k(l_i)} i_{s,t_i}^V(v), \quad (5.7)$$

the cumulative contact count up to window j :

$$\text{CumCount}_j = \sum_{i=1}^{jw} |k(l_i)|, \quad (5.8)$$

and the mean grammar rate over all encountered grammar items:

$$\text{GR}_j = \frac{1}{\sum_{i=1}^{jw} |k(l_i)|} \sum_{i=1}^{jw} \sum_{v \in k(l_i)} \text{GR}(s, t_i, v). \quad (5.9)$$

Equations (5.7), (5.8), and (5.9) summarize cumulative learning behavior over fixed-size windows. They represent the total amount of novel information encountered, the number of grammar items studied, and the average perceived proficiency in each window. This window-based aggregation smooths fluctuations and highlights broader trends in learning progression.

I then examined the relationship between these indicators and grammar proficiency using Spearman’s rank correlations, simple linear regression, and multiple linear regression.

This analysis enables us to evaluate how well the accumulation of information conceptualized by ITLAK explains learners' perceived proficiency, compared to the more straightforward metric of contact frequency. This analysis reveals how information accumulation contributes to sustained proficiency development by aggregating learning history.

(s, v) Analysis: Knowledge-Specific Effect

In the (s, v) analysis, the unit of analysis was the learner-knowledge element pair. Let the time of the end of the experimental period (i.e., the time of the last learning activity logged) be T . For each pair (s, v) , we calculated the cumulative information content for knowledge element v accumulated by learner s :

$$\text{CumInfo}_{s,v} = \sum_{\substack{l \in L_{s,T} \\ v \in k(l)}} i_{s,t_l}^V(v), \quad (5.10)$$

the cumulative contact count:

$$\text{CumCount}_{s,v} = |\{l \in L_{s,T} : v \in k(l)\}|, \quad (5.11)$$

and the final self-assessment score for the corresponding grammar item:

$$\text{GR}_{s,v} = \text{GR}(s, T, v), \quad (5.12)$$

where t_l is the time when the log l was left. These equations (5.10), (5.11), and (5.12) evaluate learning behavior for each specific grammar item. They measure how much new information the learner has accumulated, how often the item was encountered, and how well the learner believes they understand it.

I examined the relationships among these variables by computing Spearman's rank correlations. Further, we applied simple and multiple regression analyses to assess the explanatory contribution of information content compared to contact frequency. This allows us to isolate the effect of learning on each grammar item, independent of temporal sequence. By doing so, we can assess whether the perceived difficulty and learning gain are specific to particular knowledge elements, offering a complementary perspective to the temporal and behavioral analyses.

5.2.4 Result

This section shows the results of the analyses performed using the above three methods. A total of 171 learning logs were obtained during the experiment period. After manually

removing invalid data (e.g., logs with no answers or logs with system errors), 141 responses left by 18 students were deemed valid. This means that 18 unique students translated 141 EFL sentences using the proposed system.

Micro Analysis

I analyzed the level of individual learning logs ($N = 138$, 3 logs excluded from the analysis automatically). We used the mean information content $\bar{i}_v$ to examine its relation to the mean self-assessment score $GR_{s,t}$ and its explanatory power.

Spearman’s Correlations: A Spearman’s rank correlation analysis revealed a statistically significant negative correlation ($\rho = -0.242$, 95% CI $[-0.393, -0.078]$, $p < 0.01$). This suggests that learning logs associated with less frequently encountered knowledge elements (i.e., higher information content) tend to co-occur with lower self-assessment scores, which aligns with the interpretation that novelty or difficulty is reflected in lower proficiency.

Simple Regression: A simple regression model yielded a statistically significant yet small negative association between the mean information content and mean self-assessment scores (standardized $\beta = -0.08$, 95% CI $[-0.147, -0.009]$, $p < 0.05$), accounting for 2.9% of the variance ($R^2 = 0.029$). Although the explanatory power is limited, the direction and magnitude align with the preceding correlation analysis. This suggests that higher information content tends to coincide with knowledge elements learners perceive as novel or difficult. These findings support ITLAK’s theoretical claim that information content reflects perceived uncertainty in the learning experience.

Macro Analysis

Next, we analyzed learning logs aggregated by fixed-size windows of $w = 5$, thereby reflecting the cumulative learning behavior of each learner ($N = 35$). We computed each window’s cumulative information content, contact count, and average grammar proficiency score. The division into fixed-size windows was adopted to smooth out short-term fluctuations in learning behavior and capture broader trends in cumulative learning. This window size was determined by empirically adjusting the balance between granularity and data sparsity to achieve meaningful aggregation of learning logs while maintaining temporal resolution. Specifically, $w = 5$ was chosen because the number of answers per learner ranged from 1 to 22 (mean = 7.83, SD = 5.77), indicating considerable variability.

Spearman’s Correlations: Spearman’s correlation analyses at the window-aggregated level ($N = 35$) revealed weak positive association between cumulative information content and learners’ grammar self-assessment ($\rho = 0.27$, 95% CI $[-0.065, 0.557]$, $p = 0.111$) and between cumulative contact count and the same outcome ($\rho = 0.273$, 95% CI $[-0.067, 0.555]$, $p = 0.113$). Although the confidence intervals include zero and the effects are not statistically significant, the parallel direction and magnitude of the two coefficients point to a tentative positive pattern that merits further investigation with a larger sample.

Simple Regression: Using cumulative information content as the sole predictor produced a small, insignificant effect (standardized $\beta = 0.001$, 95% CI $[-0.001, 0.003]$, $p = 0.165$, $R^2 = 0.058$). Even in a model where the predictor was replaced with cumulative contact count using the same procedure, the coefficient remained at only $\beta = 0.009$ (95% CI $[-0.004, 0.022]$, $p = 0.153$, $R^2 = 0.061$). In both cases, the coefficient of variation was less than 6%, indicating that the effect size was minimal, and the results merely confirmed the weak positive trend observed in the previous Spearman’s correlation.

Multiple Regression (collinearity-adjusted model): Because the two cumulative predictors were almost perfectly correlated ($VIF \approx 798$), cumulative information content was regressed onto cumulative contact count, and the residuals were used to orthogonalize the variables. In this model, the residuals represent the part of the cumulative information content that cannot be predicted from contact count; it can be interpreted as a “novelty component”, namely, information gained from encounters that go beyond what mere repetition would provide. In the resulting model, the cumulative contact count remained positive ($\beta = 0.009$, 95% CI $[-0.004, 0.022]$, $p = 0.151$), whereas the residuals were negative ($\beta = -0.032$, 95% CI $[-0.088, 0.024]$, $p = 0.247$); neither coefficient reached significance. The model explained 10.0% of the variance in grammar self-assessment ($R^2 = 0.100$; adj. $R^2 = 0.044$; $F(2, 32) = 1.779$, $p = 0.185$), only a marginal improvement over the single-predictor regressions ($\Delta R^2 \approx 0.039 - 0.042$).

Even after removing collinearity, the ITLAK-based cumulative information count and the raw contact count offered limited explanatory power, suggesting that their effects may partly offset one another. Capturing learners’ proficiency more robustly will likely require finer-grained indicators, such as the rate of novel information exposure or explicit time-series dynamics, rather than simple cumulative counts.

(s, v) Analysis

Finally, we analyzed the level of learner-grammar item pairs (s, v) , focusing on how the accumulated learning experience with specific grammar items influenced learners' self-assessment. These grammar pairs consist of learners and grammar items each learned ($N = 617$).

Spearman's Correlations: The correlations between the cumulative indicators and proficiency were both positive and statistically significant for information content ($\rho = 0.276$, 95% CI [0.201, 0.347], $p < 0.0001$) and contact count ($\rho = 0.291$, 95% CI [0.217, 0.361], $p < 0.0001$).

Simple Regression: Cumulative information content as the sole predictor had a significant yet small explanatory effect (standardized $\beta = 0.0029$, 95% CI [0.002, 0.004], $p < 0.001$, $R^2 = 0.064$). In a model using cumulative contact count, the coefficient was slightly larger and significant (standardized $\beta = 0.0178$, 95% CI [0.012, 0.023], $p < 0.001$, $R^2 = 0.060$). Both cases had the coefficients of variation around 6%, indicating the small explanatory powers supporting this weak positive trend in the Spearman's correlations.

Multiple Regression (collinearity-adjusted model): The two raw predictor were almost perfectly collinear ($VIF \approx 140$). To disentangle their unique effects, cumulative information content was regressed on cumulative contact count, and the residuals were entered alongside the contact count. The model explained $R^2 = 0.074$ (adj. $R^2 = 0.071$; $F(2, 614) = 24.57$, $p < 0.001$), a modest but significant improvement over either single-predictor model ($\Delta R^2 \approx 0.010 - 0.014$). In other words, repetition and information-based novelty make small, independent contributions to perceived proficiency after removing their shared variance.

These findings corroborate the macro-level pattern: proficiency appears to rise with both the frequency of encounters and the novelty value of those encounters. However, each factor explains only about 6–7% of the variance, underscoring that additional variables such as learner motivation or prior knowledge are likely at play.

5.2.5 Discussion

RQ2-1: Immediate Learning Effect

The results of the micro analysis showed that proficiency scores tended to be lower for newer, i.e., less familiar, grammar items, which indicates a slight but reliable tendency

that higher information content coincides with lower self-assessment. However, the effect explains lower than 3% of the variance, so additional factors (e.g., task length, learner fatigue) are likely involved. Despite its lack of explanatory power, this tendency is consistent with the theory of Feeling of Knowing Judgement (FOK) in meta-cognition (Koriat, 1993). Metacognitive judgements regarding the memory of learned material include Ease of Learning Judgement (EOL), Judgement of Learning (JOL), and FOK (Serra & Metcalfe, 2009). Among these, FOK is a judgment that estimates whether learners know the item even when they fail to remember it. Although there were some differences between self-assessment immediately after learning and FOK, consistency was observed in that students had little sense of knowing knowledge they had encountered only a few times. In this case study, when learners were presented with English grammar items they had never learned before, namely, items with high information content, the FOK was activated, leading to the judgement that they had not understood the information, resulting in lower self-assessment scores.

RQ2-2: Cumulative Learning Effect

The macro analysis analyzed the cumulative effects of learning activities up to a specific time. The results showed that, although not significant, cumulative information content and contact frequency positively correlated with self-assessment scores. This insignificant yet positive correlation might suggest a hypothesis that simply learning English grammar extensively leads to a higher level of understanding.

However, the result that cumulative learning diminishes the effect of information content is worth observing. When comparing the same number of learning actions, the cumulative information content gained differs between repeatedly learning the same item and broadly learning a wide range of knowledge. In the former, the information content gained eventually decreases, while in the latter, the information content gained per learning session remains high. However, in actual learning situations, a wheel-spinning phenomenon (Beck & Gong, 2013) may occur, where learning outcomes are not commensurate with the information acquired. According to the result, we can hypothetically suggest that the entropy of the learner may be able to explain this situation. When entropy is high, i.e., when the learner engages in learning activities related to various types of knowledge, the value of the information acquired in a single learning session decreases. In contrast, the opposite is true when entropy is low. It suggests that the value of the

information obtained varies depending on the entropy of the learner. Since we did not compute entropy in the present study, it would be helpful to conduct analyses that take entropy into account in the future to design more effective learning support strategies.

Furthermore, the process of human learning itself is influenced by complex factors that cannot be measured by mere contact with knowledge elements, as it accumulates over time. Even considering that the explanatory power of each indicator in the simple regression analysis was not very strong, this suggests the inherent interpretive difficulty of the learning process.

RQ2-3: Knowledge-Specific Learning Effect

At the learner-knowledge (s, v) level, cumulative contact count and the informational novelty showed small but statistically reliable positive contributions to self-assessed proficiency. Although the combined model explained only 7% of the variance, the incremental R^2 over either single-predictor fit ($\Delta R^2 \approx 0.01 - 0.014$) indicates that repetition and novelty each account for a unique slice of learning progress. This complementarity suggests that effective grammar acquisition is not driven solely by “more practice” or exclusively by “more new material,” but by an interplay of the two. This implies that neither novelty nor repetition alone can fully explain perceived mastery, pointing to additional moderators such as prior knowledge or feedback quality. Therefore, regarding educational implications, it suggests that teachers must design lessons that consider increasing the number of times students learn and teach new learning items, rather than simply increasing the number of times they learn or teach new learning items alone.

Furthermore, by analyzing the knowledge element level, we must also consider the possibility that the learning load specific to each element may have been evenly distributed. For example, we can assume a case where there are differences in difficulty among items, such as those that remain challenging even after ten exposures and those that can be sufficiently understood after just three exposures. These differences in difficulty also vary among learners, and such item-level heterogeneity may have attenuated the global effect size. In future studies, it will be necessary to consider these differences in difficulty among knowledge items.

Implication

This study suggests that the information content calculated by ITLAK indicates the possibility of being an effective indicator for capturing the difficulty and novelty of the learned knowledge. In particular, from the perspective of immediate effects, which capture a single learning action, a significant negative correlation between the information content and self-assessment scores was confirmed and statistically significant, but practically small explanatory power in a simple regression analysis. In other words, when learners encountered new or challenging grammar points, they evaluated their understanding conservatively. Information content and contact frequency were significantly related to self-assessment in analyzing knowledge elements, providing insights consistent with the intuition that grammar points learned more frequently were understood better.

However, the effect of cumulative information content was not significantly different from simple contact count's, suggesting the presence of confounding factors between the two. Such confounding effects were exhibited when the accumulation was knowledge element-based. This result indicates that contact count and information content are not interchangeable but complementary indicators that capture distinct dimensions of learning: familiarity and epistemic novelty.

Overall, ITLAK could show the potential to be an effective means of capturing not only the quantity but also the quality of learning behavior. This is a theoretical contribution to KALA, particularly in interpreting short-term learning behavior and evaluating contact with new knowledge, demonstrating the usefulness of information-based analysis. However, it also became clear that the evaluation of long-term learning behavior and exposure to new knowledge was not significantly different from simple contact frequency. This may reflect the inherent difficulty of grasping the nature of learning.

If we apply these findings to the design of specific learning support, changing the support measures for short-term and long-term learning seems effective. In the short term, information-based support focusing on high-information content, i.e., new or challenging knowledge, is effective. For instance, ITLAK metrics could inform adaptive learning systems, highlighting novel or underexplored grammar items to increase engagement and reflection. In a classroom setting where learners are introduced to new grammar structures, ITLAK-based indicators can help identify which students engage with novel content and may require additional scaffolding or support. On the other hand, in the long term,

entropy-based support that measures which knowledge elements are being systematically deepened is suggested to be effective. For example, an entropy-based dashboard could visualize learners' knowledge dispersion, helping educators identify whether students are diversifying or stagnating in their learning trajectory. Although this study only suggests the possibility of entropy, such visualizations can trigger timely interventions or peer matching. Using ITLAK immediately or cumulatively can function as adequate learning support.

Limitation

We have to acknowledge the limitations of this study. First, the amount of learning data analyzed was minimal in the context of LA and cannot be considered educational big data. Therefore, the results of applying ITLAK to actual big data are still unclear in some areas. This small sample size limits generalizability; effect sizes and CIs were therefore emphasized. As a case study, the findings are illustrative, and larger-scale replication is required for generalization.

Second, the study relied on learners' subjective evaluations of English grammar items. This was due solely to technical limitations, as it is difficult to mechanically measure the level of understanding of each English grammar item. In the future, we plan to introduce objective evaluations as a comparison and conduct a comparative analysis with subjective assessment. This will improve the accuracy of ITLAK evaluations and contribute to verifying the existence of the FOK mentioned earlier.

Furthermore, the demonstration period was only five weeks, and the effects of long-term knowledge changes and forgetting were not considered. Therefore, knowledge retention, relearning, and the long-term behavior of ITLAK indicators are still unknown. However, as mentioned above, considering that the accumulation of learning suggests not only the information content but also the usefulness of entropy, it is recommended that supporting the use of more complex indicators may be helpful in the long term. Longitudinal analysis covering an entire semester or multiple interventions is necessary in the future.

Finally, it should be noted that the granularity and relevance of knowledge elements were not carefully considered. This study used the CEFR-J Grammar Profile as a list of English grammar items. Still, since this list treats each grammar item as independent, this study did not consider the semantic relationships between grammar items. Therefore, no

insights were gained into their impact on grammar item acquisition (e.g., the relationship between the past perfect subjunctive and the past perfect tense). This may affect the accuracy of information content calculations and the clarity of the relationship with self-assessment. In the future, it will be necessary to conduct empirical studies considering the sophistication of such “knowledge models” to perform a more refined analysis of ITLAK.

5.3 Study 3: OKLM Application to Learning Support

Study 3 introduces OKLM into an actual educational setting to verify the effectiveness of learning support using OKLM. In a different context from the Studies 1 and 2—specifically, an English literature reading course in higher education—we conducted an empirical study applying OKLM to a literature recommendation function. This demonstrated that learning support using OKLM can function effectively and that differences in internal features within OKLM alter the impressions formed by learners receiving such support.

5.3.1 Research Proposals and Research Questions

We proposed and formulated an OKLM, a learner modeling framework based on LA techniques, to address the aforementioned challenges in understanding learners' knowledge states. This attempt opens up the research field of KALA as LA that can deal with knowledge, such as what students have learned. Based on this work, we report a case study applying OKLM to material recommendation in EFL learning.

KALA is a new paradigm that uses data-driven methodologies to go beyond mere behavior tracking and instead focuses on LA to handle learners' knowledge. This study implies the potential of KALA by examining the effectiveness of OKLM empirically and is expected to be the first step toward opening up new possibilities for LA.

The case study aims to validate the OKLM framework by evaluating whether data-driven knowledge assessment functions effectively. Specifically, the study employs a system for recommending English academic papers in the context of undergraduate academic reading. Learners interacted with three types of OKLM-based recommendation algorithms, and we examined how these algorithms influenced learning activities and how well students accepted them. Accordingly, the research questions in this study are as follows:

RQ3-1: To what extent are knowledge and proficiency data in OKLM relevant to learning outcomes?

RQ3-2: To what extent was each of the three paper recommendation algorithms accepted by the students?

RQ3-3: To what extent did each of the three paper recommendation algorithms affect students' learning outcomes?

5.3.2 Paper Recommendation

Modeling Comprehensive Proficiency Using Knowledge Analogy

Here, we define the ease of learning and the comprehensive proficiency using the analogy of learning based on the knowledge model. This concept is based on the idea that if a learner has learned a certain knowledge concept, it will be easier to learn another concept related to the concept learned by analogy. Using relative proficiency, the easiness of learning knowledge items $el_i(S, s, v, \tau) \in [0, 1]$ is defined as follows:

$$el_i(S, s, v, \tau) = \max_{u \in V} w_{opt}(u, v) rp_i(S, s, u, \tau), \quad (5.13)$$

where $w_{opt}(u, v)$ is a relevance value between the knowledge item u and v .

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 (5.13). 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.

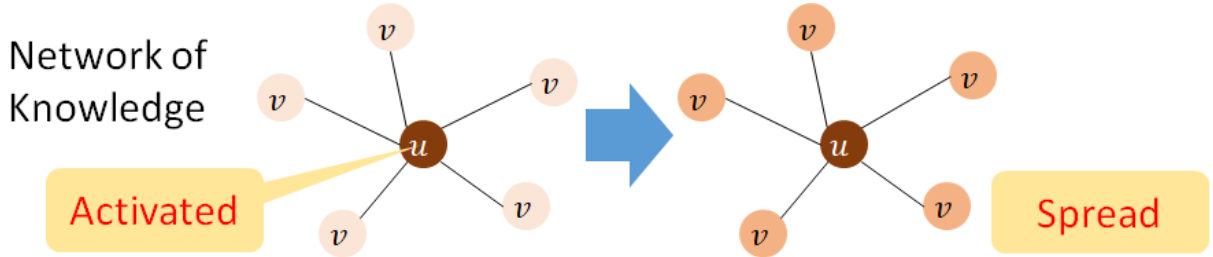


Figure 5.6: The spreading activation model of knowledge

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 5.6). 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.

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. Using the relative proficiency and the ease of learning, we can consider the value that shows how easy it is to learn a particular knowledge concept when it appears. We call it comprehensive proficiency and define it as follows:

$$cp_i(S, s, v, \tau) = c_{cp1}rp_i(S, s, v, \tau) + c_{cp2}el_i(S, s, v, \tau) \quad (5.14)$$

where c_{cp1} and c_{cp2} are constants, $c_{cp1} \geq 0$, $c_{cp2} \geq 0$, and $c_{cp1} + c_{cp2} = 1$.

Recommendation Algorithm

In this study, we adopted a method for material recommendation using relative and comprehensive proficiency. Before introducing the algorithm, we assume that a learning material can be represented as a set of knowledge concepts included in it. We can define V_m as a set of knowledge concepts included in the material m . However, which materials are recommended depends on which knowledge items they contain. Therefore, it makes sense to calculate a recommendation score for each knowledge item and to use their sum as the material's score.

First, we assumed that materials that contain much knowledge that a learner has not mastered well before should be recommended because learning these materials will broaden the learner's knowledge and help the learner overcome his/her difficulties. In this case, we can use the value of the function cp_i as an index for scoring the knowledge contained in the materials. Based on this idea, we defined a recommendation weight denoted as f_i^{cp} , which takes a higher value as the material is strongly recommended. In other words, the recommendation algorithm highly rates knowledge in a material that they have not yet mastered.

$$f_i^{cp}(S, s, v, \tau) = 1 - cp_i(S, s, v, \tau). \quad (5.15)$$

Next, we considered the time elapsed since the learner last learned a particular knowledge concept. Once the knowledge is learned and not reviewed, it will be forgotten as time passes. Therefore, we believe it is meaningful to encourage learners not to forget the knowledge they have learned by reminding them using the recommendation algorithm. In the following equation, $\tau_{last}^{i,s,v}$ refers to the time when the learner s last took a learning

action for a knowledge concept v before the time τ .

$$f_i^{org}(s, v, \tau) = 1 - \frac{1}{c_{org}(\tau - \tau_{last}^{i,s,v}) + 1}. \quad (5.16)$$

The third method considers other learners' knowledge. In an environment where learners are learning collaboratively, the knowledge that other learners have learned well is probably worth learning, and it is important to understand what they are learning. This method gives high ratings to knowledge concepts peer learners can master well. Using the other learners' information, the recommendation weight can be formulated as follows:

$$f_i^{sig}(S, s, v, \tau) = \frac{1}{|S| - 1} \sum_{s' \in S, s' \neq s} rp_i(S, s', u, \tau). \quad (5.17)$$

Synthesizing these methods, the recommendation score of the material m for the learner s can be defined as follows. All the above methods can be combined: we can determine which perspective of recommendations to stress in the recommendation.

$$F_i(S, s, v, \tau) = \alpha f_i^{cp}(S, s, v, \tau) + \beta f_i^{org}(s, v, \tau) + \gamma f_i^{sig}(S, s, v, \tau), \quad (5.18)$$

$$\text{score}_i(S, s, m, \tau) = \frac{1}{|V_m||V|} \sum_{u \in V_m} \sum_{v \in V} w_{opt}(u, v) F_i(S, s, v, \tau). \quad (5.19)$$

5.3.3 Setting

Participants

The participants of this study were undergraduate students enrolled in a lecture "A Study of Foreign Literature" at a Japanese national University. This lecture aimed to nurture students' reading skills in English academic literature through activities such as reading, summarizing, and reviewing English papers. A total of 45 students registered for the course, of which 30 completed the requirements for the final assessment.

Experimental Settings

This lecture included activities for students to read, summarize, and review English academic papers. We selected 65 papers from Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK19).

The lecture was conducted over nine weeks (Weeks 6-14 of a 14-week course), divided into three tasks: Weeks 6-8, 9-11, and 12-14. In each task, students were required to

select one article for reading and write a summary and a reflection. The reflection included a logical analysis of the article’s structure, organization, and potential flaws and an introspective component addressing difficult sections or relevance to their own research interests.

Each task followed a three-stage structure. First, students selected papers using the proposed paper recommendation system, read them, and wrote a reflection paper. Although allowed to read multiple papers, they submit a reflection paper about one paper. Next, they exchanged reflections with a peer, received comments on their strengths and areas for revision, and submitted a revised version. Finally, in group discussions, students shared their reading and reflections, created a PowerPoint summary, and delivered a group presentation about their reflection papers.

In this lecture, students used a paper recommendation system (Figure 5.7) developed for this study to select articles. This recommendation system shows students all the papers in order of recommendation rating. The rating was calculated using the formula (5.19), multiplied by 100 for clarity, and then output. From the system’s recommended list, each student chose one paper. The recommendation criteria of the system varied across tasks (Table 5.12), allowing analysis of how different algorithms affected learner behavior and group outcomes.

Recommendation with OKLM			説明
1.Exploring the Subtleties of Agency and Indirect Control in Digital Lea	推薦度:	37	
2.The Impact of an Online Tutoring Program for Algebra Readiness on Math	推薦度:	36.9	
3.Reliable Deep Grade Prediction with Uncertainty Estimation	推薦度:	36.3	
4.DiAd- Domain Adaptation for Learning at Scale	推薦度:	36.2	
5.user2code2vec- Embeddings for Profiling Students Based on Distribution	推薦度:	36.1	
6.Differences in Student Trajectories via Filtered Time Series Analysis	推薦度:	35.8	
7.Exploring students' sensemaking of learning analytics dashboards- Does	推薦度:	35.8	
8.Where You Are, Not What You See- The Impact of Learning Environment on	推薦度:	35.4	
9.DEBE feedback for large lecture classroom analytics	推薦度:	35.4	
10.Introducing meaning to clicks- Towards traced-measures of self-efficac	推薦度:	35	

Figure 5.7: Screenshot of a paper recommendation system. The higher up the list, the more strongly recommended the paper is.

The recommendation conditions were changed by changing the values of the variables α , β , and γ in the formula 5.19. Table 5.12 summarizes these variables’ settings and the algorithms’ corresponding purposes. The parameters in the above formulae were set as follows: $c_{rp1} = 0.7$, $c_{rp2} = 0.3$, and $c_{forg} = 1$.

Table 5.12: Parameter settings and algorithmic focus in each recommendation condition

Task	α : Less familiar vocabulary	β : Forgotten vocabulary	γ : Common vocabulary among others	Primary focus of the algorithm
1	0.8	0.1	0.1	Recommend papers that include words unfamiliar to the learner.
2	0.1	0.8	0.1	Recommend papers that include words the learner may be forgetting.
3	0.1	0.1	0.8	Recommend papers that include words commonly learned by others.

Knowledge Model Generation

This study built a knowledge model of English vocabulary using the following procedure:

1. Identify 50 English words that are common nouns of 3 letters or more from the material.
2. Vectorize the 50 words using word2vec (Mikolov et al., 2013).
3. Build a minimum spanning tree using the cosine similarity between those 50 vectors.
4. The resulting tree structure is used as the knowledge model, with the new weights being the weights of the minimum spanning tree subtracted from 1.

In step 2, we used a pre-trained model word vectors for Simple English trained on Wikipedia (Bojanowski et al., 2017). Step 3 uses a minimum spanning tree with cosine similarity between the vectors of the 50 words obtained. Therefore, the branches in the resulting tree structure have smaller weights for words connected by words that are close in meaning. In Step 4, the weights of the branches are modified so that the branches connecting words that are closer in meaning have larger weights.

5.3.4 Data Collection

Data Collection Platform

The OKLM was also prepared in parallel to calculate the students' English proficiency. The OKLM contained 65 papers prepared for this course, and the English vocabulary knowledge models were extracted from each paper. A data flow was also prepared in which the BookRoll usage logs stored in the LRS are sent to OKLM. Since the students were instructed to use the marker functions primarily during the course (see the next section), the OKLM updated the proficiency level of each word highlighted with markers in BookRoll.

The paper recommendation system was implemented in LogPalette (Majumdar et al., 2022), a tool for analyzing BookRoll usage logs. LogPalette retrieved the paper recommendations generated by the OKLM using its proficiency data and displayed them to the students.

Students' Learning Activities

During the intervention, students were instructed to read papers using BookRoll. As part of the reading activity (the first stage of the experiment), they were asked to use BookRoll's marker function to highlight specific vocabulary items while reading. Two types of markers were used to indicate different cognitive responses to words: yellow markers for words that the student did not understand and red markers for words that the student considered important. These marking activities were part of the reading strategy and a data source for modeling learner knowledge in the OKLM. Specifically, the vocabulary items marked during reading were treated as knowledge elements, and students' interactions with them contributed to the estimation of proficiency levels within OKLM.

Each time a student marked a word using BookRoll, a learning log was generated and stored in OKLM. The system updated the proficiency value (i.e., $x_i(s, v, \tau)$) of the corresponding knowledge element (i.e., the English word) according to the following rule:

- When a word was marked for the first time, a new log was recorded with a proficiency value of 0.05.
- If the word had already been marked, the new log was assigned a proficiency value equal to the most recent value multiplied by 1.05.

- All learning logs generated from these marking actions were labeled with the indicator type “Apply” (i.e., $i = \text{Apply}$), denoting active engagement with the vocabulary.

Through this process, the OKLM could incrementally construct individualized learner models based on observed reading behaviors, using the accumulation and scaling of learning logs to estimate proficiency for each encountered vocabulary item. In the proposed paper recommendation, relative proficiency $rp_i(S, s, v, \tau)$ (the inter-learner deviation of the value of $x_i(s, v, \tau)$) and comprehensive proficiency $cp_i(S, s, v, \tau)$ calculated from it, were used (see equations (5.15) to (5.19)). In this case, the elements of the set S were all students who attended the class.

Assignments

Each student was required to submit one written assignment per task, resulting in three assignments corresponding to the three recommendation periods. The assignment consisted of an English reflection paper in which students responded critically to one of the academic papers they had read during the task. Hereinafter, reflection papers will be referred to as reports.

The course instructor evaluated and graded all submitted reports based on content quality and language use. In addition to the score assigned, two linguistic features of the reports were extracted for subsequent analysis: word count (the total number of words contained in each report) and vocabulary size (the number of unique words used). These two features served as quantitative indicators of productive language use and were used as dependent variables in later analyses related to learning outcomes and algorithm effects.

Questionnaire

Questionnaires were administered to capture students’ perceptions of the paper recommendation system and their learning experiences throughout the study. These questionnaires served as the primary data source for addressing RQ3-2 and RQ3-3. Two types of questionnaire items were included:

- RQ3-2-related items assessed the overall acceptability and usefulness of the paper recommendation system. These items asked students to express their agreement (on a 5-point Likert scale) with statements such as:

1. Reading academic papers through the recommendation system was a valuable learning activity.
 2. They would like to engage in similar activities in the future.
 3. They are willing to use this system again actively.
- RQ3-3-related items asked students to reflect on the differences among the three recommendation conditions and to indicate which algorithm best matched the intended effects of the recommendation system. Specifically, students were asked to select the recommendation period during which (1) the recommended paper best matched their English proficiency, (2) the paper was most interesting to them, and (3) they encountered the most significant number of unknown vocabulary items.

Questionnaires were administered after the entire study (post-questionnaire). In the subsequent analysis, we primarily used data from the post-questionnaire, as it provided students with a basis for comparing all three recommendation conditions. We also provided a free response section at the end of the questionnaire. The complete list of questionnaire items is provided in the Appendix (Table 1).

5.3.5 Analysis Method

RQ3-1: To what extent are knowledge and proficiency data in OKLM relevant to learning outcomes?

To address RQ3-1, we examined the correlation between learner data recorded in the OKLM and four distinct indicators of learning outcomes during each recommendation period. From the OKLM logs, we derived two types of indicators for each of the three recommendation periods. These indicators reflect the extent and depth of knowledge engagement and acquisition during each learning phase.

- **Knowledge coverage:** The number of distinct English vocabulary items (knowledge elements) encountered by each student through the recommended papers.
- **Average proficiency:** The mean of estimated proficiency values across all knowledge elements computed by the OKLM.

Learning outcomes were assessed using the following four variables, each aggregated over the same three recommendation periods. Collectively, these indicators reflect both behavioral engagement and productive language proficiency.

1. **Class participation score**, assigned by the course instructor.
2. **Report score**, also assigned by the instructor.
3. **Report word count**, based on the total number of words used in each report.
4. **Report vocabulary size**, defined as the number of unique vocabulary items used.

I analyzed Spearman correlation coefficients between the two OKLM-derived indicators and the four learning outcome variables. Eight correlations were analyzed per recommendation period. Along with the correlation coefficients, p-values were calculated to assess statistical significance. This analysis was conducted separately for each recommendation condition, allowing us to investigate whether the relationship between knowledge modeling and learning outcomes varies depending on the recommendation strategy.

By analyzing the magnitude and direction of correlations between OKLM data and concrete learning outputs, we evaluated whether the system’s internal representations of student knowledge and proficiency are meaningfully associated with observable learning outcomes. This provides empirical insight into the validity of OKLM’s modeling approach in capturing learning progress.

RQ3-2: To what extent was each of the three paper recommendation algorithms accepted by the students?

To address RQ3-2, we analyzed both system usage logs and responses to the post-experiment questionnaire. From the system usage log data of the paper recommendation system, the following four behavioral indicators were extracted:

- **Display count**: The number of times recommendation results were displayed.
- **Click count**: The number of times a recommended paper was clicked.
- **Acceptance rate**: Calculated as (Click count) / (Display count), representing the proportion of recommended papers clicked. This value can exceed one since more than one paper can be clicked on the system displayed at once.
- **Number of selected papers**: The total number of papers students selected before finalizing their choice.

These indicators were compared across the three recommendation algorithm conditions using Friedman tests, followed by post-hoc Nemenyi tests when significant differences were observed. The non-parametric Friedman test was chosen due to the repeated-measures design and non-normal distribution of some indicators.

In addition to behavioral data, we collected students' answers to the post-experiment questionnaire. The questionnaire responses provided complementary insights into their subjective impressions and preferences on the recommendation system, thus enabling a more comprehensive evaluation of algorithmic acceptance.

RQ3-3: To what extent did each of the three paper recommendation algorithms affect students' learning outcomes?

To address RQ3, we analyzed the quality of student-generated output (written reports) and their subjective evaluations of the learning impact. After engaging with each of the three recommendation algorithms, students were required to submit an English-written report corresponding to each paper-reading session (three reports per student in total). We evaluated the variables for each report: report scores, word count, and vocabulary size. These three variables were treated as objective indicators of students' learning outcomes. Since each student produced three reports under different recommendation conditions (within-subject design), we employed Friedman tests to examine statistical differences across the three algorithm groups, followed by Nemenyi post-hoc tests where appropriate.

To supplement the objective analysis, we included three self-report items in the post-experiment questionnaire, asking students to identify which recommendation algorithm:

- Best matched their English proficiency level.
- Offered the most interesting papers.
- Provided the most significant exposure to unknown vocabulary.

Responses were aggregated and visualized as proportions to assess the perceived effectiveness of each algorithm in facilitating learning. Although these items were not subjected to statistical testing, they provided valuable contextual insight into student-perceived learning effects.

Evaluation of written reports served as a direct measure of learning outcomes, reflecting both productive vocabulary use and instructor-judged academic performance.

Meanwhile, the questionnaire responses helped clarify how students interpreted each algorithm’s educational value in relevance, interest, and challenge. These sources enabled a multifaceted analysis of how the different recommendation algorithms may have supported student learning.

5.3.6 Result

Correlation between OKLM Indicators and Learning Outcomes

To answer RQ3-1, we examined the relationship between learner data generated in OKLM and four learning outcome variables: class participation score, report score, report word count, and report vocabulary size. Spearman correlation coefficients were computed separately for each recommendation period (Task 1-3), using two OKLM-derived indicators: knowledge coverage and average proficiency.

In Task 1, as shown in Table 5.13, both OKLM indicators showed a significant positive correlation with class participation: knowledge coverage ($r = .42, p = .022$), and average proficiency ($r = .52, p = .003$). These results suggest that students who encountered more vocabulary and achieved higher proficiency levels tended to participate more actively in class during Task 1. However, no statistically significant correlations were found between OKLM indicators and report-related outcomes (report score, wordcount, vocabulary size).

In Tasks 2 and 3, no significant correlations were observed between OKLM indicators and any learning outcome variables. All correlation coefficients were weak ($r < .28$) and non-significant ($p > .13$), suggesting limited associations between modeled knowledge or proficiency and observable learning products in these later periods.

Perceived and Behavioral Acceptance of Recommendation Algorithms

To answer RQ3-2, we compared learner behaviors across the three recommendation conditions (Task 1-3) using four system-derived behavioral indicators: Display count, Click count, Acceptance rate, and Number of selected papers. All indicators were compared using Friedman tests, and where significant differences were found, Nemenyi post-hoc tests were conducted. The results are summarized below.

Display count: The number of times the recommendation list was displayed differed significantly across tasks ($\chi^2(2) = 19.0, p < .001$). Task 2 again showed significantly lower

Table 5.13: Correlation coefficients between OKLM indicators and learning outcomes

Task	OKLM Indicator	Class Participation	Report Score	Report Word Count	Report Vocabulary Size
1	Knowledge coverage	.42 ($p = .022$)	.03 ($p = .894$)	.11 ($p = .564$)	.07 ($p = .720$)
	Avg. proficiency	.53 ($p = .003$)	.11 ($p = .555$)	.15 ($p = .439$)	.15 ($p = .440$)
2	Knowledge coverage	.25 ($p = .182$)	.21 ($p = .258$)	.04 ($p = .830$)	.12 ($p = .529$)
	Avg. proficiency	.24 ($p = .199$)	.26 ($p = .160$)	.06 ($p = .750$)	.14 ($p = .462$)
3	Knowledge coverage	.28 ($p = .137$)	.06 ($p = .744$)	.01 ($p = .940$)	.05 ($p = .781$)
	Avg. proficiency	.28 ($p = .137$)	.06 ($p = .744$)	.01 ($p = .940$)	.05 ($p = .781$)

Note. Bold values indicate $p < .05$.

values than both Task 1 ($p < .001$) and Task 3 ($p < .05$), indicating a similar pattern of lower interaction (Figure 5.8a).

Click count: A Friedman test also revealed a significant difference in the number of clicks among the three conditions ($\chi^2(2) = 12.9, p < .01$). Post-hoc tests showed that Task 2 resulted in significantly fewer clicks than both Task 1 ($p < .01$) and Task 3 ($p < .05$), suggesting reduced engagement with recommendations in that period (Figure 5.8b).

Acceptance rate: No significant difference in acceptance rate was found among the three conditions ($\chi^2(2) = 5.69, p = .06$), indicating that the likelihood of clicking a recommended paper when presented was relatively stable regardless of the algorithm used (Figure 5.8c).

Number of selected papers: Although visual inspection suggests possible differences (Figure 5.8d), statistical testing did not reveal any significant differences across tasks

($\chi^2(2) = 5.26, p = .07$). Therefore, no further post-hoc analysis was conducted for this variable.

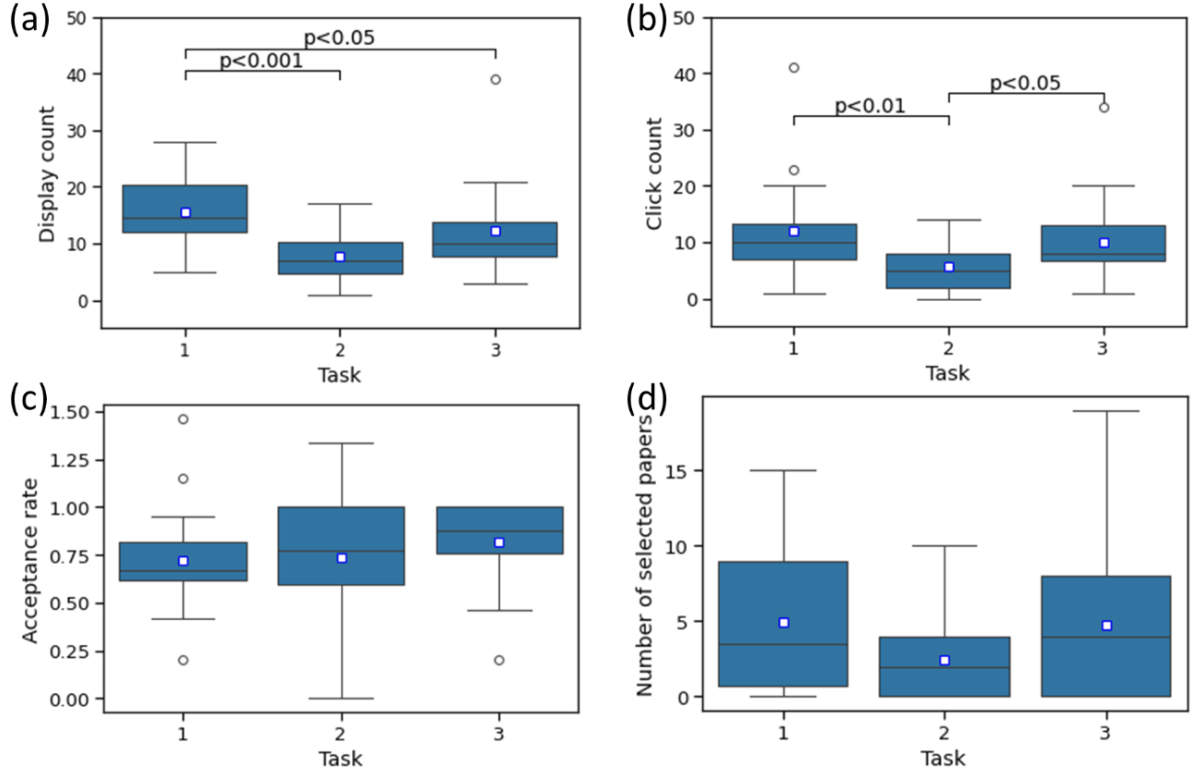


Figure 5.8: (a) Number of recommendation displays per task, (b) Number of clicks on recommended papers across the three tasks, (c) Acceptance rate (clicks/displays) per task, and (d) Number of papers selected before final decision.

To complement the behavioral data, we analyzed students' responses to the post-experiment questionnaire, which assessed their subjective acceptance of the paper recommendation system. Participants rated the three statements in the questionnaire using a 5-point Likert scale (1=Stringly disagree, 5=Strongly agree). Figure 5.9 shows the distribution of responses for each item using violin plots. The first item received the most consistently high ratings, indicating that students generally recognized the value of the activity. In contrast the second and third items, which assessed students' willingness to repeat or actively engage with the system, showed greater variation, with some students expressing hesitation or neutrality. This suggests a possible gap between perceived educational value and behavioral intention to reuse the system. While students acknowledged the activity as beneficial, they may not have been equally enthusiastic about repeating it

or using it proactively in the future.

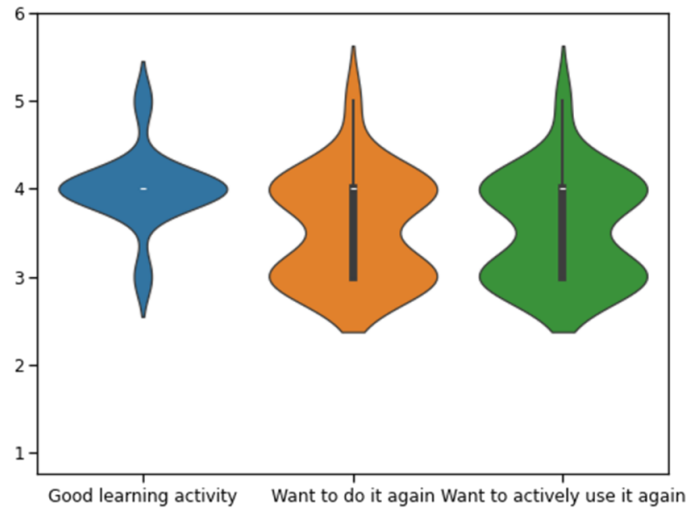


Figure 5.9: Violin plots of students' responses to subjective evaluation items (1-5 Likert scale)

In the free response section of the questionnaire, two comments were referring to the possible usefulness of the system in the literature review in the graduation research, and two comments about reading papers of interest. In addition, there was one comment that they wanted to use the system not only to find papers in their field of interest, but also to search for papers in new fields that would be useful for their research and broaden their horizons. This suggests that the paper recommendation system can be helpful in the classroom setting of literature reading, academic research, and self-study.

Behavioural Engagement across Recommendation Conditions

To address RQ3-3, we compared students' learning outcomes across the three recommendation conditions using the three variables: report score, word count, and vocabulary size. A Friedman test was conducted for each variable, followed by Nemenyi post-hoc tests where appropriate. The results are summarized below.

Report score: No significant difference was observed in instructor-assigned report scores across the three recommendation periods ($\chi^2(2) = 2.25, p = .11$). As shown in Figure 5.10a, the median scores remained relatively consistent among all tasks.

Report word count: The number of words in students' reports did not differ significantly between conditions ($\chi^2(c) = 5.80, p = .06$), as shown in Figure 5.10b. Although Task 2 shows a slight increase in media, the variation within groups was substantial, and the difference was not statistically meaningful.

Report vocabulary size: A significant difference was found in the number of unique words across tasks ($\chi^2(2) = 6.82, p < .05$). Post-hoc comparisons revealed that Task 1 reports contained significantly fewer unique words than those in Task 3 ($p < .05$), as shown in Figure 5.10c. No other pairwise comparisons were significant.

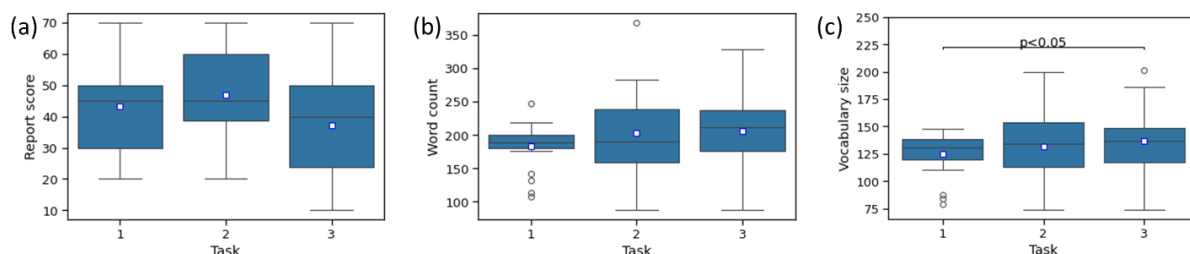


Figure 5.10: Comparison of report-based learning outcomes across three recommendation conditions. (a) Report score, (b) Total word count, and (c) Vocabulary size

This suggests that the recommendation algorithm used in Task 3 may have supported broader vocabulary usage in students' written outputs. However, no consistent effects were observed for report length or instructor-evaluated quality.

To further investigate the impact of the recommendation algorithms on students' learning outcomes, we analyzed responses to three items in the post-experiment questionnaire. Each of the items asked students to select one of the three recommendation tasks (1 to 3) that they felt best fulfilled a particular effect on (1) suitability to the English proficiency, (2) recommendation of interesting papers, and (3) exposure to unfamiliar vocabulary.

The distribution of responses are shown in Figure 5.11. Although none of the items showed statistically significant differences ($p > .05$, χ^2 test), the trends offer complementary insights to the report-based outcome measures discussed earlier. In particular, Task 1 was selected most frequently in item (1), suggesting that students perceived it as the most vocabulary-enriching, consistent with the higher vocabulary size observed in Task reports. Conversely, Task 2 was most frequently chosen in items (2) and (3), suggesting that students found its recommendations more aligned with their proficiency and

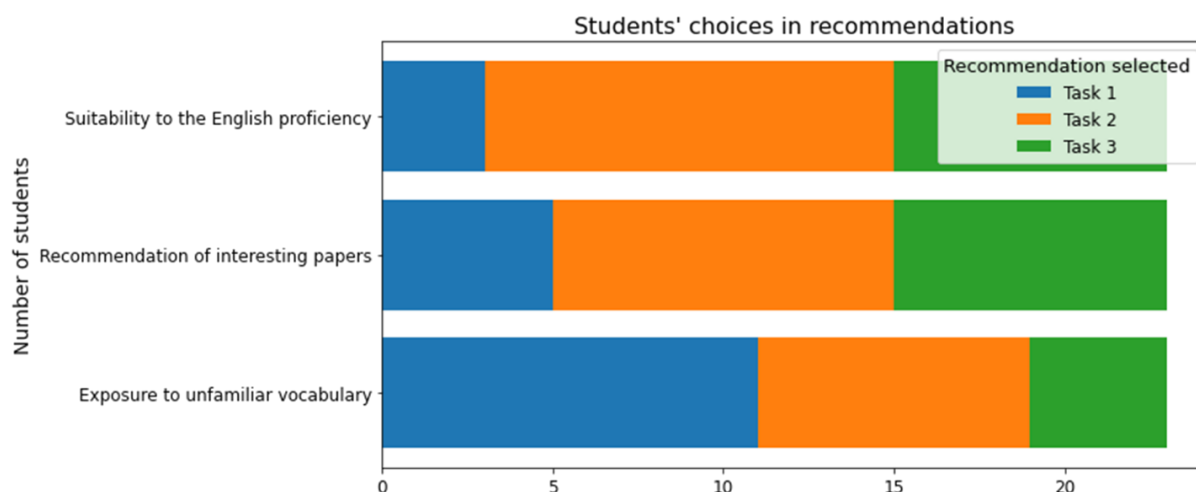


Figure 5.11: Distribution of responses to three subjective evaluation items

interests. This possibly explains their improved textual output despite lower behavioral engagement.

5.3.7 Discussion and Implications

RQ3-1: To what extent are knowledge and proficiency data in OKLM relevant to learning outcomes?

Only in Task 1 was a significant positive correlation found between OKLM indicators (number of knowledge elements and average proficiency) and learners' class participation scores. This may reflect learners' higher initial motivation and shared starting point, as many were unfamiliar with academic English reading. Active classroom engagement likely translated into more frequent marking and vocabulary interaction, which OKLM captured.

The OKLM's current proficiency model, assigning 0.05 for a first marker and multiplying subsequent values by 1.05, may have amplified early-stage behaviors, making early engagement more visible in Task 1. This interpretation aligns with post-questionnaire results indicating that Task 1 exposed learners to the most unfamiliar vocabulary.

However, this result should be interpreted with caution. Task 1 coincided with learners' first exposure to the system and the activity structure, potentially generating a novelty effect that learners are more likely to actively engage with systems that are new to them (Sung & Mayer, 2013). Additionally, the technical nature of the reading materials such as frequently used technical terms may have inflated perceived and modeled difficulty.

No significant correlations were found in Tasks 2 and 3. Possible explanations include diminished novelty, increased routinization or automation of behavior (Deci & Ryan, 2000; Winne & Hadwin, 1998), and the OKLM’s limited sensitivity to gradual, mid-course learning. Since its proficiency model grows exponentially, smaller reinforcement behaviors may not be adequately reflected. Thus future implementations could address this by incorporating time-based decay functions or frequency-based logs (e.g., Corbett and Anderson, 1994; Pardos and Heffernan, 2010). It may allow for more nuanced modeling across early and sustained learning phases.

RQ3-2: To what extent was each of the three paper recommendation algorithms accepted by the students?

Learners’ behavioral engagement with the system, measured by the number of displays, clicks, and selections, declined over time (Task 1 > Task 3 > Task 2), likely reflecting a novelty effect in Task (Sung & Mayer, 2013). Interesting, subjective evaluations from the post-questionnaire diverged from this trend: Task 2 received the highest ratings regarding interest and alignment with learners’ proficiency. This discrepancy highlights the distinction between perceived usefulness and actual usage, as discussed in prior research (Kizilcec et al., 2013; F. Li et al., 2023).

The number of reviewed papers before selection followed the same declining trend (Task 1 > Task 3 > Task 2), suggesting that learners initially explored more options, later becoming more efficient or guided in their choices. This may reflect that students used the system merely because it was mandatory. However, since the subjective evaluation shows that students had the most opportunities to encounter unfamiliar words in Task 1, we consider that the algorithm in Task 1 could achieve its objectives, to recommend papers that include word unfamiliar to learners. On the other hand, this feature might have increased the learners’ burden, which our future research should tackle with.

Overall, each recommendation algorithm elicited distinct responses:

- Task 1, which emphasized unfamiliar words, led to high interaction but low satisfaction: likely due to increased difficulty.
- Task 2, which targeted forgotten vocabulary, saw lower usage but higher satisfaction, possibly due to learners’ recognition of familiar words and a sense of mastery (Loewenstein, 1994).

- Task 3, focused on peer-common vocabulary, resulted in moderate interaction and feedback, potentially due to reduced personalization.

Despite recognizing the educational value of the activity, learners’ willingness to repeat or actively engage with the system remained neutral, indicating that appreciation alone does not guarantee continued use (F. Li et al., 2023). Open-ended responses, however, suggested that learners saw potential beyond classroom use, such as for research or independent exploration, highlighting both instrumental and intrinsic motivations. Thus, limitations of the study include that no questionnaire item explicitly asked which recommendation was most appropriate, leaving some ambiguity in interpreting preference. Supporting these dual needs may be critical to sustaining long-term system engagement.

RQ3-3: To what extent did each of the three paper recommendation algorithms affect students’ learning outcomes?

Only vocabulary size showed a significant difference between Task 1 and Task 3 among all outcome variables. The Task 3 algorithm, which recommended papers containing common vocabulary shared by peers, may have supported learners’ productive use of broadly applicable vocabulary in writing (Nation, 2006). Although not the algorithm’s primary goal, its emphasis on high-frequency words likely facilitated broader lexical output. However, OKLM indicators in Task 3 showed no significant correlation with report data, likely because learners rarely marked common words and thus were underrepresented in the system. This points to a design limitation in OKLM, which infers knowledge only from explicitly marked items. Words familiar to learners may be central to writing but invisible to the model, highlighting the need for more inclusive trace mechanisms.

Repeated exposure to report tasks may also explain the observed vocabulary growth. Task repetition has been shown to reinforce domain-specific vocabulary through natural usage (Ellis, 2005). Working on reports repeatedly can also help students naturally learn the specific vocabulary, because the papers covered in class were quite specialized in the LA research field.

The recommendations in Task 3 were based on peer data, raising personalization and modeling validity concerns. Common vocabulary among various learners did not provide individual recommendations for each learner. Marker frequency alone may overlook contextually important but rare words, which should have been considered important in the

document. TF-IDF or embedding-based approaches (Griffiths & Steyvers, 2004; Takii, Flanagan, et al., 2024) could help surface domain-relevant vocabulary.

Interestingly, despite receiving the highest subjective ratings, Task 2 did not produce measurable learning gains. This means that the recommendation in Task 2 was rated as interesting and suitable for students’ skills, but this impression was not reflected in the learning outcomes. This may reflect the gap between receptive and productive vocabulary knowledge (Laufer, 1998). Although the vocabulary that formed the basis for the recommendations in Task 2 was somewhat familiar to the learners, they had not sufficiently mastered it, so it was difficult for them to reflect on their learning output. This implies that learners may recognize forgotten words but not actively use them in writing (Waring & Takaki, 2003).

Finally, the study relied on quantitative measures (e.g., word count, vocabulary size, report score), which overlook qualitative aspects such as coherence or critical thinking. Future work should incorporate richer linguistic or discourse-level analyses.

Overall, the impact of recommendation strategies is shaped by complex interactions between vocabulary type, learner readiness, and system design. Statistical differences alone may not capture this interplay; a nuanced, context-sensitive perspective is needed.

Theoretical Implications

The primary theoretical contribution of this study is based on the proposal of OKLM: a paradigm that shifts the focus of LA from observable behaviors to structured representations of knowledge. Building on critiques that LA has remained agnostic to what learners are learning (Tsai & Gasevic, 2017; Wilson et al., 2017), we introduced the OKLM as a proof-of-concept that quantifies learner interactions with vocabulary knowledge and enables content-based recommendation. This approach advances beyond behaviorist or black-box modeling by allowing semantically interpretable, content-level analytics of learner behaviors. Our findings demonstrate that knowledge-aware interventions can address epistemic goals such as reinforcing forgotten knowledge or encouraging exposure to new concepts rather than merely tracking engagement or correctness. At the same time, observed discrepancies between OKLM logs and learners’ vocabulary use in writing highlight the limitations of trace-based inference, even in knowledge-aware systems. Thus, OKLM offers a technical model for knowledge-level analytics and calls for a reexamination of what constitutes valid evidence in data-driven learning research.

Practical Implications

The findings also offer actionable implications for educational practice and system design. Learners expressed interest in using OKLM-based recommendations beyond classwork (e.g., literature searches for thesis writing or exploring unfamiliar domains), highlighting its potential for autonomous learning. Moreover, varying recommendation criteria (e.g., unfamiliar, forgotten, or commonly known vocabulary) enabled distinct learner experiences, suggesting the value of adaptive interfaces that align algorithmic logic with learner goals and contexts. However, the current OKLM design, which relies on explicit learner-generated markers, may overlook important but familiar vocabulary. Future implementations could integrate content-sensitive importance metrics such as TF-IDF or domain-based embedding models (Griffiths & Steyvers, 2004) and refine proficiency modeling using time-sensitive mechanisms like knowledge decay (Pardos & Heffernan, 2010). Overall, the effectiveness of knowledge-aware systems depends on algorithmic design and how learners interact with these systems within specific educational contexts. This study lays the groundwork for context-aware, goal-aligned LA systems that balance technical sophistication with pedagogical relevance.

Limitation

This study has several limitations that should inform future research. First, OKLM relies on learner-generated markers, which may overlook familiar or low-salience vocabulary. This likely contributed to the mismatch observed in Task 3, where learners' vocabulary use increased despite limited OKLM updates. Incorporating implicit behavioral traces (e.g., gaze or scrolling) or content-sensitive metrics (e.g., TF-IDF) may enhance coverage and accuracy. Additionally, the current multiplicative model for proficiency accumulation emphasizes early behavior but may not capture reinforcement or decay. Adopting time-aware models or spaced repetition principles could better reflect the temporal dynamics of learning.

Second, the study assessed outcomes using quantitative indicators (report scores, word counts, vocabulary size), which omit qualitative dimensions such as coherence or argumentation. Future work should adopt mixed-methods approaches, including rubric-based or NLP-assisted analysis of learner writing. The absence of a control group also limits the ability to isolate the effects of recommendation strategies, and learners' unawareness of

algorithmic differences leaves the role of transparency unexplored. Finally, the study was conducted in a single course at one institution, so findings may not generalize to other settings. Broader validation of the OKLM framework will require diverse educational contexts, content domains, and learner populations.

5.4 Summary of the Three Studies

5.4.1 Satisfaction of Requirements in Each Study

This section describes how each study satisfies the OKLM system requirements, specifically the three types of openness. Table 5.14 summarizes the results of this analysis.

Table 5.14: Requirement fulfillment of OKLM’s three opennesses

Study	Knowledge-domain	Learner	LA System
Study 1	✓		
Study 2	✓		✓
Study 3	✓		✓
Visualization	✓	✓	✓

Study 1 modeled vocabulary knowledge for English extensive reading materials. This modeling approach employed the vocabulary knowledge map construction method proposed by Flanagan, Chen, et al. (2019). This method is a general approach for organizing English vocabulary into a graph structure and does not depend on OKLM characteristics. Therefore, at this stage, OKLM’s knowledge-domain openness can be considered satisfied. However, since the English material recommendation system, as a learning support system, did not employ LA methodology, LA system openness remains unsatisfied. Furthermore, as it lacks functionality to disclose knowledge status to users, learner openness is also not satisfied.

Similar to Study 1, Study 2 also targeted Japanese junior high school students. This was an empirical experiment using English-to-Japanese translation exercises, focusing on modeling English grammar knowledge. Here, a method was employed to organize CEFR-J English grammar elements into a graph structure, which does not depend on OKLM. Therefore, knowledge-domain openness is satisfied. Furthermore, as a learner modeling approach, the methodology of ITLAK was adopted to calculate the amount of information learners acquire through learning actions. This demonstrated that the foundation for estimating learners’ knowledge states can shift from simple frequency of exposure to knowledge to criteria based on more detailed information, and that this approach can function effectively.

While Studies 1 and 2 targeted Japanese junior high school students, Study 3 is an

empirical experiment conducted in an English academic paper reading class for university undergraduates. Although no progress was observed in meeting the three types of openness requirements, this study demonstrated LA system openness by capturing detailed behavioral logs such as note taking and highlighting. Regarding learner openness, it is expected to be fulfilled by implementing the aforementioned OKLM visualization system in the classroom and making it available to users.

To summarize the above:

- Knowledge modeling in OKLM was demonstrated to be feasible for both vocabulary and grammar—core elements of language learning—albeit within the limited domain of English learning. Thus, knowledge-domain openness was satisfied.
- Learner openness is expected to be guaranteed in the future by implementing the OKLM visualization functionality in the field.
- LA system openness was satisfied by enabling the detailed learning behaviors captured by the LEAF system to serve as the basis for representing knowledge states.

Chapter 6

General Discussion

6.1 Theoretical Implications

This thesis makes a series of theoretical contributions to the field of LA by proposing the OKLM and its associated formalism, the ITLAK framework. Together, these developments address long-standing theoretical challenges in LA: how to systematically represent what learners know, how to link that representation to learning behavior data, and how to quantify the informational value of learning activities.

6.1.1 Reconceptualizing Learner Modeling in LA

Traditional LA approaches have primarily focused on analyzing behavioral traces, such as page views, highlighting, or quiz responses, because they are easily captured and quantified. While this has enabled large-scale monitoring, it has also limited LA’s explanatory power regarding knowledge acquisition itself. Learner models developed in adjacent domains (e.g., Knowledge space theory, BKT, DKT) have provided more sophisticated representations of learner states, but they typically rely on predefined knowledge structures and assessment responses. Consequently, they remain difficult to generalize across contexts and to integrate with the wide variety of LA systems now in use.

The OKLM framework reconceptualizes this problem by establishing a mapping between learning logs and knowledge components. Rather than presupposing a closed, static domain, OKLM advocates for an “open” approach: the openness for knowledge-domain, learner, and LA system. These three dimensions of openness reframe learner modeling as a dynamic, adaptable, and integrative process that overcomes the rigidity of previous frameworks. Theoretically, this positions OKLM as a generalizable paradigm for KALA.

6.1.2 Formalizing Information in Learning

A second theoretical implication lies in the formalization of the informational value of learning activities. Through ITLAK, the thesis demonstrates that each learning log can be associated with a measure of information content, derived from information theory. This represents a conceptual shift: instead of merely counting encounters with knowledge elements, LA can treat each encounter as carrying a quantifiable informational weight, reflecting novelty, predictability, and uncertainty. This attempt to theorize the information content of learning will not only philosophically reexamine the value of learning but also open new avenues for elucidating its objective and cognitive mechanisms. It holds the potential to truly establish educational informatics, which has been merely an applied science of information technology, as the informatics of education in the literal sense.

Empirical findings illustrate the nuanced effects of information content: individual logs with high novelty may temporarily reduce self-assessed comprehension, whereas cumulative information accrual positively correlates with learners' sense of progress. This duality highlights a theoretical principle of learning as both a process of momentary disruption (novelty, surprise) and long-term consolidation (accumulated knowledge). Such a principle provides a bridge between cognitive theories of learning (e.g., the role of desirable difficulties, interleaving, and spacing) and computational modeling in LA.

6.1.3 A Network-Based Knowledge Perspective

The OKLM framework also advances a network-based view of knowledge representation. By constructing knowledge graphs from learning materials and connecting them with observed behaviors, OKLM situates LA within a broader theoretical shift toward relational, graph-based epistemologies. Knowledge is not merely an inventory of items but a structured set of interdependent concepts, where learning behavior traverses and reshapes the network. This perspective aligns with contemporary approaches in cognitive science and knowledge representation, while offering a computationally tractable model for LA.

Theoretically, such a perspective opens new possibilities for investigating pathways of learning, optimal sequencing of knowledge components, and cross-domain transfer. It suggests that LA should not be confined to “snapshots” of learner proficiency, but rather should model the topology of knowledge acquisition and its evolution over time.

6.1.4 Toward a Unified Theory of KALA

By unifying behavioral traces, knowledge representations, and information-theoretic measures within a single framework, this thesis advances the conceptual foundations of LA. The theoretical contribution lies not only in the introduction of OKLM and ITLAK as specific frameworks, but also in the articulation of a more general vision: a KALA. KALA emphasizes that understanding and supporting learning requires moving from “learning context analysis” toward “knowledge analysis,” thereby enabling LA to more directly address its original aspiration of understanding and optimizing learning.

In sum, the theoretical implications of this work can be summarized in three points:

1. Generalizability of learner modeling through the three openness principles of OKLM.
2. Quantification of informational value of learning actions via ITLAK, offering a principled link between data and theory.
3. A network-based epistemology of learning, situating LA in alignment with contemporary knowledge sciences.

Together, these implications establish a foundation for future research that treats LA not only as a tool for practical optimization, but also as a scientific discipline capable of theorizing learning itself.

6.2 Pedagogical Implications

While the theoretical implications of this thesis establish a foundation for KALA, the pedagogical implications highlight how the OKLM and the ITLAK framework can be applied in real educational settings. The experimental studies presented in this thesis not only demonstrate feasibility but also suggest concrete practices for enhancing teaching, learning, and educational system design.

6.2.1 Explainable Recommendations for Learners and Teachers

One of the clearest practical contributions lies in the development of explainable recommendation systems. By grounding book and paper recommendations in knowledge models, the proposed systems allow learners to understand why a particular resource is suggested. This transparency increases trust and acceptance, as shown in both extensive

reading and academic reading contexts. For teachers, explainability serves as a pedagogical tool: it enables them to guide learners by referring to the underlying knowledge elements rather than relying on opaque algorithmic outputs. Such systems exemplify how LA can evolve from predictive dashboards to interpretable decision-support mechanisms, aligning with broader calls for ethical and trustworthy AI in education.

6.2.2 Novelty and Repetition as a Design Pattern

The second study demonstrated that the information-theoretic view of learning clarifies the dual roles of novelty and repetition. While novelty introduces cognitive challenges that may initially reduce perceived comprehension, repetition consolidates learning and increases self-efficacy over time. This suggests a practical instructional design pattern: “Novelty $\times$ Repetition.” Educators and instructional designers can deliberately combine surprising, unfamiliar knowledge encounters with repeated exposures to reinforce understanding. OKLM provides the computational means to balance these two forces, allowing for adaptive curriculum sequencing that is both engaging and effective.

6.2.3 Classroom Management and Formative Assessment

By mapping learning logs onto knowledge components, OKLM enables teachers to visualize learners’ weaknesses in fine-grained ways. For instance, dashboards informed by OKLM can highlight specific knowledge elements where learners struggle, rather than reporting only aggregate performance. This supports formative assessment practices, enabling timely intervention and personalized feedback. Moreover, such knowledge-tagged visualization tools can empower learners themselves to self-regulate by recognizing gaps in their knowledge networks, thereby fostering metacognition and learner autonomy.

6.2.4 Evaluation Indicators for Educational Effectiveness

The thesis also introduces new indicators for evaluating learning processes: knowledge coverage, mean proficiency, and information content. These measures extend beyond conventional metrics such as scores or time-on-task. They allow instructors and system developers to evaluate learning outcomes in terms of how comprehensively a learner has traversed the knowledge space, how deeply they have consolidated specific elements, and how much informational novelty was encountered during learning. These indicators pro-

vide actionable data for curriculum refinement and system design, offering a new standard for evaluating educational interventions in knowledge-intensive subjects.

6.2.5 Scalability and Cross-Context Application

Finally, OKLM’s three dimensions of openness for knowledge-domain, learner, and LA system ensure scalability across diverse contexts. Once a knowledge model is constructed, the framework can be applied to multiple subjects (e.g., language learning, mathematics, or science). Its openness to learners and teachers enhances usability across educational levels, while its interoperability with other LA systems allows for institutional or even cross-institutional deployment. The case studies in this thesis illustrate how OKLM can be integrated into both school-based programs and higher education environments, suggesting its potential as an infrastructure for large-scale, cross-curricular, and cross-school knowledge-aware analytics.

6.2.6 Toward Evidence-Informed Educational Practice

Taken together, these practical contributions emphasize that LA can move beyond retrospective analysis and static dashboards. With OKLM and ITLAK, educators gain tools that are both interpretable and actionable, learners benefit from tailored recommendations and self-regulation support, and institutions can adopt scalable frameworks for data-driven decision-making. Ultimately, the practical implication of this thesis is to demonstrate how knowledge-aware LA can bridge the gap between theoretical insight and classroom reality, thereby supporting a more evidence-informed and learner-centered educational practice.

6.3 Limitations and Future Work

Despite its theoretical and practical contributions, this thesis also has limitations that should be acknowledged. These limitations point toward promising directions for future research.

First, the accuracy of the knowledge models remains an important constraint. Although the proposed framework allows knowledge-domain openness, the quality of subsequent learner modeling depends on the completeness and validity of the constructed

knowledge models. Domain-specific biases or inconsistencies in model construction may affect the precision of knowledge mapping.

Second, the external validity of evaluation metrics such as knowledge coverage, mean proficiency, and information content has yet to be fully established. While these indicators demonstrated utility in the presented case studies, further testing across subjects and populations is necessary to ensure their robustness and interpretability.

Third, the empirical studies were conducted with a limited sample size. Although results suggest meaningful trends, larger-scale deployments are required to validate statistical findings, especially regarding the correlation between information content and learners’ self-assessments.

Building on these limitations, three avenues for future research can be highlighted.

1. **Theoretical deepening:** Future studies should refine the KALA framework by further exploring what unique insights it can generate compared to conventional LA. This may include formal investigations of optimal sequencing, knowledge propagation, and inter-domain transfer.
2. **Practical expansion:** Beyond experimental studies, the framework should be extended toward real-world educational ecosystems. In particular, the transition from “learner models” to digital twins of learners holds potential for richer simulation, prediction, and personalization of learning experiences.
3. **Cross-context validation:** Applying OKLM across diverse subjects, institutions, and cultural contexts will be essential to assess its generalizability. Such work would also help to build repositories of open knowledge models that can serve as shared resources for the LA community.

Chapter 7

Conclusions

This thesis aimed to integrate the knowledge dimension into the field of LA, which had primarily focused on learners' behavioral aspects. While much existing LA literature has centered on behavioral traces and learning characteristics, this study argued that such data alone is insufficient to capture the essence of learning. It also argues that even if learner knowledge can be captured, it must be in an interpretable form that enables interoperability across multiple systems. Therefore, by introducing OKLM and ITLAK, this study provides the foundation for KALA. Here, learning actions are interpreted not merely as behaviors, but as events carrying informational value related to knowledge elements.

This study demonstrated that OKLM can bridge learning logs and knowledge representations by systematically mapping observable actions to knowledge elements. OKLM was shown to possess cross-context adaptability, transparency to learners and educators, and interoperability with other LA systems through its three openness principles: knowledge-domain, learner, and LA system. This flexibility represents a conceptual advancement over static, domain-specific learner models, enabling a more dynamic understanding of what learners acquire and how they construct knowledge networks.

Complementing this framework, ITLAK expanded the scope of LA by formalizing the concept of information content in learning. Empirical analysis revealed that novelty in learning often reduces short-term self-assessment, while cumulative information accumulation promotes long-term progress recognition. These findings reveal the dual nature of learning as both disruptive and integrative, linking LA with cognitive and educational theories concerning desirable difficulty, iteration, and metacognitive awareness. Thus, this thesis not only provides a methodological framework but also offers theoretical insights

into the mechanisms by which learners interact with knowledge.

Empirical studies conducted in the contexts of ER, IR, and academic reading substantiated the usefulness of OKLM and ITLAK. Explainable recommendation systems demonstrated that transparency enhances both learner trust and teacher guidance. Identifying the “dual action of novelty and repetition” as a design principle provides concrete guidance for instructional design, while knowledge-based metrics such as coverage, mastery, and information load were introduced as new standards for evaluating learning effectiveness. These applications demonstrate that the proposed framework moves LA beyond post-monitoring, deepening it into actionable and interpretable support for educational practice.

The contributions of this research can be summarized in three points:

1. Identifying three openness principles—knowledge-domain, learner, and LA system—as requirements for designing knowledge models that represent learner knowledge using LAs, and proposing, designing, developing, and implementing OKLM, a learner model embodying these principles.
2. Through OKLM, proposing the concept of KALA, an LA that handles knowledge.
3. Using OKLM, demonstrating the effectiveness of a method for objectively and quantitatively representing learner knowledge using LA.

In conclusion, this thesis positions LAs not merely as data-driven tools for optimization, but as a scientific field capable of theorizing the very essence of learning itself. By shifting from “learning context analysis” to “knowledge analysis,” OKLM and ITLAK provide conceptual and methodological resources for next-generation research and practice. The findings of this research highlight the potential of LA to reconcile rigor with humanity. Specifically, it is an LA that respects the fundamental complexity of the learning process while providing concrete and interpretable support to learners and educators. Although limitations remain in the precision of knowledge models, the generalizability of metrics, and the scale of empirical validation, the direction indicated by this research opens a path for future studies that combine a theoretical foundation with practical impact.

Publications

Journal Papers (Peer-Reviewed)

1. Takii, K., Flanagan, B., Li, H., Yang, Y., Koike, K., & Ogata, H. (2025). Explainable eBook recommendation for extensive reading in K-12 EFL learning. *Research and Practice in Technology Enhanced Learning*, 20, 027.
2. Takii, K., Liang, C., & Ogata, H. (2025). Information as Interpretation: Measuring Learning Behavior for Knowledge Insight. *IEEE Access*, Volume 13, pp.124197-124210.
3. Takii, K., Liang, C., & Ogata, H. Representing What Was Learned: An Open Knowledge and Learner Model and Its Implementation in Paper Recommendation. *Journal of Educational Technology and Society*. (under review).

International Conferences (Peer-Reviewed)

1. Takii, K., Flanagan, B. & Ogata, H. (2020), Efficiency or Engagement: Two Book Recommendation Approaches in English Extensive Reading, In *28th International Conference on Computers in Education*, Vol.2, (pp.106-112).
2. Takii, K., Flanagan, B., & Ogata, H. (2021). An English picture-book recommender system for extensive reading using vocabulary knowledge map. In *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge* (pp. 40-42).
3. Takii, K., Flanagan, B., & Ogata, H. (2021). Efl vocabulary learning using a learning analytics-based e-book and recommender platform. In *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp. 254-256). IEEE.

4. Takii, K., Flanagan, B., & Ogata, H. (2022). Explainable English Material Recommendation Using an Information Retrieval Technique for EFL Learning. In *30th International Conference on Computers in Education, Vol.1* (pp.561-570).
5. Takii, K., Tanimura, N., Flanagan, B., & Ogata, H. (2023). Construction of an English Grammar Quiz Recommendation System Using Explanation by a Knowledge Map. In *31th International Conference on Computers in Education*.
6. Takii, K., Koike, K., Horikoshi, I., Flanagan, B., & Ogata, H. (2024). OKLM: A universal Learner model integrating everyday learning activities with knowledge Maps. In *Companion Proceedings of the 14th International Conference on Learning Analytics and Knowledge* (pp. 191-193).
7. Takii, K., Liang, C., & Ogata, H. (2024). Open Knowledge and Learner Model: Mathematical Representation and Applications as Learning Support Foundation in EFL. In *32th International Conference on Computers in Education, Vol.1*, (pp.595-604).
8. Takii, K., Liang, C., & Ogata, H. (2025). Understanding Learner Behavior Using Information Theory on Learning Analytics and Knowledge. In *International Conference on Learning Evidence and Analytics*.

References

- Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11), 1–37.
- Advanced Distributed Learning Initiative (ADL). (n.d.). Experience api (xapi) [Accessed: 2025-05-02].
- Aikawa, N., Maeda, S., Mogi, T., Koike, K., Tomoto, T., Imai, I., Horiguchi, T., & Hirashima, T. (2024). Classroom implementation of an auxiliary problem presentation system for mechanics adapted to learners' errors. *Research & Practice in Technology Enhanced Learning*, 19.
- Ain, Q. U., Chatti, M. A., Meteng Kamdem, P. A., Alatrash, R., Joarder, S., & Siepmann, C. (2024). Learner modeling and recommendation of learning resources using personal knowledge graphs. *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 273–283.
- Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16*, 431–440.
- Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. *Proceedings of the 8th international conference on learning analytics and knowledge*, 41–50.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the association for computational linguistics*, 5, 135–146.
- Brantmeier, C. (2005). Nonlinguistic variables in advanced second language reading: Learners' self-assessment and enjoyment. *Foreign Language Annals*, 38(4), 494–504.
- Bull, S. (2020). There are open learner models about! *IEEE Transactions on Learning Technologies*, 13(2), 425–448.
- Chen, W. (2019). Knowledge-aware learning analytics for smart learning. *Procedia Computer Science*, 159, 1957–1965.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, 82(6), 407.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2), 240–247.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4, 253–278.

- Council of Europe. (2025). Common european framework of reference for languages [Accessed: 2025-04-30]. <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227–268.
- Dillenbourg, P., & Self, J. (1992). A framework for learner modelling. *Interactive learning environments*, 2(2), 111–137.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, 27(2), 141–172.
- 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., Chen, M.-R. A., Lecailliez, L., Majumdar, R., Akçapınar, G., Ocheja, P., & Ogata, H. (2019). Automatic vocabulary study map generation by semantic context and learning material analysis. *International Conference on Computers in Education*.
- Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge map creation for modeling learning behaviors in digital learning environments. *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)*, 428–436.
- Goto Butler, Y., & Lee, J. (2010). The effects of self-assessment among young learners of english. *Language testing*, 27(1), 5–31.
- Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology*, 50(5), 2572–2593.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National academy of Sciences*, 101(suppl_1), 5228–5235.
- Gunasekara, S., & Saarela, M. (2025). Explainable ai in education: Techniques and qualitative assessment. *Applied Sciences*, (3).
- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.
- Hooshyar, D., Pedaste, M., Saks, K., Leijen, Ä., Bardone, E., & Wang, M. (2020). Open learner models in supporting self-regulated learning in higher education: A systematic literature review. *Computers & education*, 154, 103878.
- Hwang, G.-J. (2014). Definition, framework and research issues of smart learning environments—a context-aware ubiquitous learning perspective. *Smart Learning Environments*, 1, 1–14.
- Jayaprakash, S. M., Moody, E. W., Lauria, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6–47.
- Kasim, U., & Raisha, S. (2017). Efl students’ reading comprehension problems: Linguistic and non-linguistic complexities. *English Education Journal (EEJ)*, 8(3), 308–321.
- Khadir, A. C., Aliane, H., & Guessoum, A. (2021). Ontology learning: Grand tour and challenges. *Computer Science Review*, 39, 100339.

- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the third international conference on learning analytics and knowledge*, 170–179.
- Knight, S., & Buckingham Shum, S. (2017). Theory and learning analytics. *Handbook of learning analytics*, 1, 17–22.
- Koriat, A. (1993). How do we know that we know? the accessibility model of the feeling of knowing. *Psychological review*, 100(4), 609.
- Krathwohl, D. R. (2002). A revision of bloom’s taxonomy: An overview. *Theory into practice*, 41(4), 212–218.
- Lang, C., Wise, A. F., Merceron, A., Gašević, D., & Siemens, G. (2022). What is learning analytics. *The handbook of learning analytics*, 8–18.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied linguistics*, 19(2), 255–271.
- Li, F., Zhang, H., Yan, H., Chu, H., Zhang, D., & Wong, C. U. I. (2023). Subjective behavior and technological perception: Exploring university students’ engagement intention in online gamified learning. *Proceedings of the 7th International Conference on Education and Multimedia Technology*, 243–251.
- Li, H., Majumdar, R., Chen, M.-R. A., & Ogata, H. (2021). Goal-oriented active learning (goal) system to promote reading engagement, self-directed learning behavior, and motivation in extensive reading. *Computers & Education*, 171, 104239.
- Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1), 75.
- Majumdar, R., Liang, C., Kuromiya, H., Li, H., Flanagan, B., & Ogata, H. (2022). Learning and evidence analytics framework (leaf): Innovating log data driven services for teaching and learning. *Interactive demo ISLS2022*, 24–27.
- Majumdar, R., Yang, Y. Y., Akçapinar, G., Flanagan, B., Ogata, H., et al. (2018). Goal: Supporting learner’s development of self-direction skills using health and learning data. *International Conference on Computers in Education*.
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of memory and language*, 52(4), 463–477.
- MEXT. (2020). The image of the transformation of learning brought by “1 device for 1 student with a high-speed network” [Online; accessed 2025-09-11]. https://www.mext.go.jp/en/content/20200716-mxt_kokusai-000005414_04.pdf
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Moodle.org. (2017). Moodle - open-source learning platform [Accessed: 2025-04-30]. <https://moodle.org/>
- Nakamoto, R. (2025). Development of a mathematics learning support system through the analysis of handwritten responses and self-explanations.
- Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Canadian modern language review*, 63(1), 59–82.
- Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review*, 4(4), 251–277.

- Ocheja, P., Flanagan, B., Ueda, H., & Ogata, H. (2019). Managing lifelong learning records through blockchain. *Research and Practice in Technology Enhanced Learning*, 14(1), 1–19.
- Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., & Takii, K. (2024). EX-AIT: Educational eXplainable artificial intelligent tools for personalized learning. *Research and Practice in Technology Enhanced Learning*, 19.
- Ogata, H., Majumdar, R., Flanagan, B., & Kuromiya, H. (2024). Learning analytics and evidence-based k12 education in japan: Usage of data-driven services for mobile learning across two years. *International Journal of Mobile Learning and Organisation*, 18(1), 15–48.
- OpenAI. (2025). Chatgpt: Language model developed by openai [Accessed: 2025-05-02].
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4), 49–64.
- Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a bayesian networks implementation of knowledge tracing. *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings 18*, 255–266.
- Park, S. Y., Nam, M.-W., & Cha, S.-B. (2012). University students’ behavioral intention to use mobile learning: Evaluating the technology acceptance model. *British journal of educational technology*, 43(4), 592–605.
- Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: Opportunities and challenges. *Artificial intelligence review*, 56(11), 13071–13102.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in neural information processing systems*, 28.
- Qrquez, M., & Ab Rashid, R. (2017). Reading comprehension difficulties among efl learners: The case of first and second year students at yarmouk university in jordan. *Arab World English Journal (AWEJ) Volume*, 8.
- Qu, K., Li, K. C., Wong, B. T., Wu, M. M., & Liu, M. (2024). A survey of knowledge graph approaches and applications in education. *Electronics*, 13(13), 2537.
- Rajaraman, A., & Ullman, J. D. (2011). *Mining of massive datasets*. Autoedicion.
- Sarac, S., Enisa, M., & Akgun, E. (2022). Does e-learning trigger epistemic curiosity? *Journal of Qualitative Research in Education*, (30).
- Satriani, E. (2018). Reading comprehension difficulties encountered by english students of islamic university of riau. *J-SHMIC: Journal of English for Academic*, 5(2), 15–26.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The modern language journal*, 95(1), 26–43.
- Serra, M. J., & Metcalfe, J. (2009). Effective implementation of metacognition. *Handbook of metacognition in education* (pp. 278–298). Routledge.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379–423.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400.

- SoLAR. (2025). What is learning analytics - society for learning analytics research (solar) [Online; accessed 2025-09-11]. <https://www.solaresearch.org/about/what-is-learning-analytics/>
- Sung, E., & Mayer, R. E. (2013). Online multimedia learning with mobile devices and desktop computers: An experimental test of clark's methods-not-media hypothesis. *Computers in Human Behavior*, 29(3), 639–647.
- Takami, K., Flanagan, B., Dai, Y., & Ogata, H. (2023). Personality-based tailored explainable recommendation for trustworthy smart learning system in the age of artificial intelligence. *Smart Learning Environments*, 10(1), 65.
- Takii, K., Flanagan, B., Li, H., Yang, Y., Koike, K., & Ogata, H. (2024). Explainable ebook recommendation for extensive reading in k-12 efl learning. *Research and Practice in Technology Enhanced Learning*, 20.
- Takii, K., Koike, K., Horikoshi, I., Flanagan, B., & Ogata, H. (2024). Oklm: A universal learner model integrating everyday learning activities with knowledge maps. *Companion Proceedings 14th International Conference on Learning Analytics & Knowledge, Japan*, 191–193.
- Takii, K., Liang, C., & Ogata, H. (2024). Open knowledge and learner model: Mathematical representation and applications as learning support foundation in EFL. *International Conference on Computers in Education*.
- Takii, K., Liang, C., & Ogata, H. (2025). Information as interpretation: Measuring learning behavior for knowledge insight. *IEEE Access*.
- Tono, Y. (2020). The CEFR-J wordlist version 1.6 [Accessed: 2025-04-30]. <http://www.cefr-j.org/download.html>
- Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. *Proceedings of the seventh international learning analytics & knowledge conference*, 233–242.
- UNESCO. (2015). *Rethinking education: Towards a global common good?* United Nations Educational, Scientific; Cultural Organization. <https://doi.org/10.54675/MDZL5552>
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509.
- Wang, J., Flanagan, B., & Ogata, H. (2017). Semi-automatic construction of ontology based on data mining technique. *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 511–515.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2).
- Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education*, 22(8), 991–1007.
- Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning.
- Wu, H., Wang, Z., Wang, K., Omran, P. G., & Li, J. (2023). Rule learning over knowledge graphs: A review. *Transactions on Graph Data and Knowledge*, 1(1), 7–1.
- Yang, Y., Majumdar, R., Li, H., Flanagan, B., & Ogata, H. (2024). Design of a learning dashboard to enhance reading outcomes and self-directed learning behaviors in out-of-class extensive reading. *Interactive Learning Environments*, 32(3), 892–909.

- Zanellati, A., Di Mitri, D., Gabbrielli, M., & Levrini, O. (2024). Hybrid models for knowledge tracing: A systematic literature review. *IEEE Transactions on Learning Technologies*, 17, 1021–1036.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2), 64–70.

Appendices

Table 1: Complete list of the post-questionnaire items.

Related RQ	Question
3-2	Using the paper recommendation system to read papers is a good thing.
3-2	I want to engage in reading papers using the paper recommendation system again.
3-2	I want to use the paper recommendation system actively.
3-3	Among the three times you used the paper recommendation system, which paper recommendation was most suitable for your English skills?
3-3	Among the three times you used the paper recommendation system, which paper recommendation system recommended the most interesting papers to you?
3-3	Among the three times you used the paper recommendation system, which paper recommendation exposed you to the most unfamiliar words?
None (free-text)	Please write your thoughts on how you would like to use the paper recommendation in other forms or other purposes.