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

Proficiency Modeling in Junior High Math: Adapted Cognitive Statistical Models to E-Book Learning Contexts

Changhao LIANG[a*], Kensuke TAKII[b] & Hiroaki OGATA[a]
[a]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
[b]*Graduate School of Social Informatics, Kyoto University, Japan*
*liang.changhao.8h@kyoto-u.ac.jp

Abstract: Digital learning platforms equipped with behavior sensors have provided abundant educational data. Utilizing this data, learner modeling can identify assorted learner characteristics from their behavior logs for learning analytics and dynamically update them in real time. There is a growing demand for knowledge-level modeling, moving beyond behavioral logs to assess knowledge proficiency. Based on item response theories and cognitive statistical models, existing studies estimate learning rates across various knowledge elements in each learning step in intelligent tutoring systems. However, these models, tailored to specific knowledge domains, offer limited flexibility across different knowledge units and scenarios. This paper introduces an adaptation of the basic additive factor model underpinned by logistic regression, focusing on behavior indicators. Drawing upon authentic learning data from an e-book learning infrastructure for junior high math, we examine the feasibility of our adapted models and demonstrate their potential for flexibility across knowledge units and learning phases.

Keywords: additive factor model, cognitive statistical model, learner model, junior high math, e-book learning, learning analytics

1. Introduction

Learner models in digital learning platforms cover both domain-specific and domain-independent information, quantified as learning evidence (Boticki et al., 2019). These indicators come from platforms equipped with data sensors, such as e-book reading logs, test scores, and other pertinent data, which can dynamically update with logs. Among these models, knowledge holds vital importance, as the primary goal of online education is knowledge acquisition. This underscores the importance of understanding learners' knowledge states (Fischer et al., 2002), especially in K-12 contexts where knowledge levels are predominantly assessed from a cognitive perspective. Therefore, interventions driven by knowledge-based data are emphasized in such settings.

In platforms like e-books, students engage with learning material through markers and memos, leaving various recorded behaviors. Research has explored how such engagement can be modeled and visualized for educational predictions using learning logs (Chen et al., 2021). However, these outputs lack **interpretability regarding knowledge states**, addressing behaviors in content levels without clear correspondence to knowledge profiling. In mastery learning contexts such as math, active engagement does not necessarily guarantee success, as wheel spinning may occur (Wang et al., 2020). Therefore, it is imperative to reconsider how to effectively model learners' knowledge proficiency based on various types of e-book records.

On the other hand, cognitive statistical models, originally applied for intelligent tutoring systems (ITS; Draney et al., 2012), incorporate dynamic interactions and dialog for each learning step. Logistic regression models are popular for knowledge modeling, providing a “probability” measure. Assuming each target problem is a knowledge unit, the probability of a learner successfully solving a problem can serve as a promising indicator of knowledge proficiency (Chounta et al., 2017).

Extending from conventional item response models and underpinned by logistic regression, the Additive Factor Model (AFM) considers prior learning opportunities (Cen et al., 2006). It is applicable for various digital learning platforms holding these records. Existing breakthroughs in cognitive models focus on weighting conjunctive knowledge components in learning steps, considering difficulty, slipping rate, and learning rate for each piece of knowledge (Chi et al., 2011; MacLellan et al., 2015). However, these coefficients are domain-specific, lacking **flexibility across contexts**. Macro elements, such as general learning behaviors, show promise, although fewer cognitive models inspected learning rates for general behavioral interactions and broader contexts.

Inspired by the basic AFM model and its variations, which consider learning rate coefficients of prior learning opportunities, this paper proposes an adapted instructional factor analysis model for knowledge proficiency based on an e-book context. Instead of estimating learning rates for respective knowledge or problems, we focus on general behavior indicators, for adaptability on broader learning scenarios. Drawing upon authentic learning data from junior high math, we investigated the feasibility of our adapted models and demonstrated their potential for flexibility across scenarios.

2. Research foundations

2.1 OKLM conceptualization from Learning analytics infrastructure

With the advancement of learning analytics (LA), integrated learning platforms such as the Learning and Evidence Analytics Framework (LEAF) enable the collection of everyday learning logs (Ogata et al., 2024b). These logs, stored at the behavior level as primary learner models, provide a foundation for LA in learning predictions. One key challenge in developing such infrastructure is integrating various logs across learning resources to make it open-ended (Bull, 2020). To address this, the conceptualization of the Open Knowledge and Learner Model (OKLM) was proposed (Takii et al., 2024). This model aims to integrate scattered behavior logs into a knowledge level, so as to facilitate interoperability across learning systems and contexts.

Figure 1 shows a conceptual diagram of the learner model construction in OKLM. OKLM presents a learner model capable of managing and tracking which knowledge items are covered by each learning action, linking the learner's daily learning logs to a knowledge map generated from learning resources such as e-books. Each node in the knowledge map contains information about the learning logs for its corresponding resources. To capture the complex nature of knowledge acquisition and proficiency, OKLM connects the verbs in xAPI logs from behavior sensors to Bloom's taxonomy that represents multifaceted cognitive development (see Figure 2; Krathwohl, 2002).

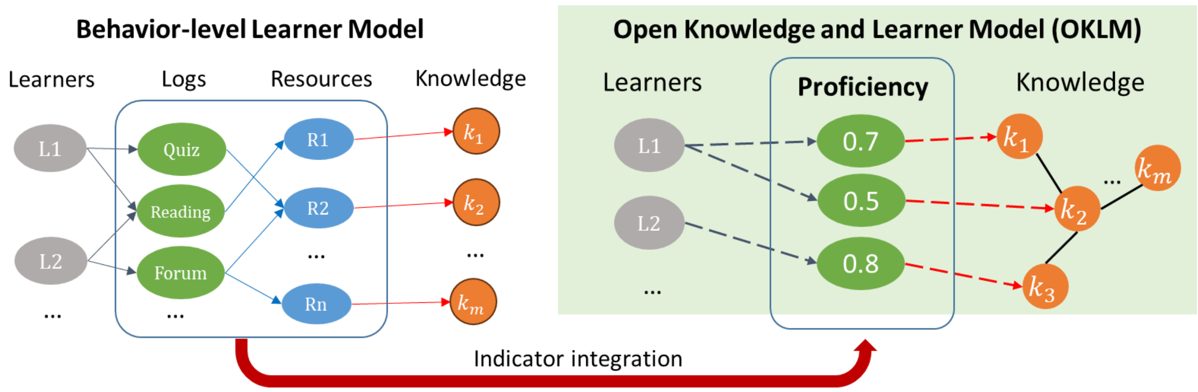

Figure 1. The conceptual diagram of the learner model construction in OKLM.

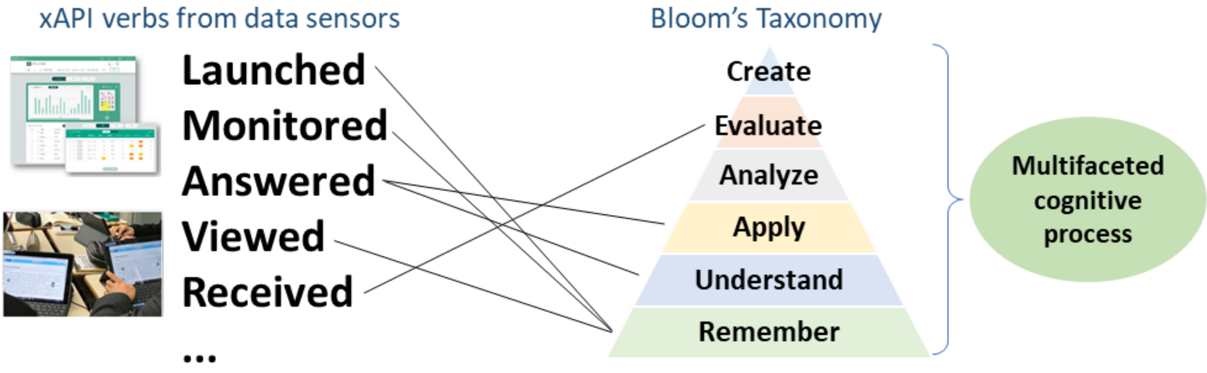

Figure 2. Example connection of xAPI verbs and the layers of Bloom's Taxonomy.

Based on these connections constructed from educational big data in digital learning platforms, OKLM offers the potential to support learning, including learning status visualization, material recommendation, and grade prediction (Khadir et al., 2021; Takii et al., 2024). The model aids in integrating data from different learning tools based on connected knowledge units. Beyond conventional data-driven systems, applications based on OKLM design can provide explainable interventions based on learners' knowledge structures, rather than relying solely on black box models or simple log counts (Ogata et al., 2024a; Shin, 2021).

To realize these applications, a fundamental challenge is to create Learner-Knowledge (L-K) connections from daily Learner-Resource (L-R) logs and Knowledge-Resource (K-R) labels assigned by instructors. Thus, the OKLM conception lays the conceptual foundation for bridging the gap of knowledge proficiency profiling from raw behavior logs.

2.2 Cognitive modeling in intelligent tutoring systems

In cognitive learning, learners progress through multiple learning steps, including Q&A dialogs, quizzes, or peer assistance, to acquire knowledge (Draney et al., 2012). To estimate the knowledge accumulation process from these steps, the frequency of previous exposure to relevant knowledge units is commonly considered.

The Additive Factor Model (AFM; Cen et al., 2006) and its variations, originally extending from item response theory (Draney et al., 2012), are popular in cognitive modeling for predicting performance at each learning step. Due to its logistic model properties, the output of AFM models can quantify learners' knowledge state, as predicted probability values can indicate proficiency on the target learning goal as a percentage. This logistic model also aligns with the Zone of Proximal Development (ZPD) theory (Vygotsky, 1980), as the "gray areas" near the classification threshold of predicted probability can represent such zones for learning recommendations (Chounta et al., 2017), thereby holding further educational significance.

$$z_i = \alpha_i + \sum_{kc} \left(\beta_{kc} + \gamma_{kc} \times N(kc, i)\right) \tag{1}$$

$$z_i = \alpha_i + \sum_{kc} \left(\beta_{kc} + \mu_{kc} \times S(kc, i) + \rho_{kc} \times F(kc, i)\right) \tag{2}$$

$$z_i = \alpha_i + \sum_{kc} \left(\beta_{kc} + \mu_{kc} \times S(kc, i) + \rho_{kc} \times F(kc, i) + \nu_{kc} \times T(kc, i)\right) \tag{3}$$

The formulas of three popular AFM models from Chi et al. (2011) are depicted above, with the predicted probability $p_i = 1/(1 + z_i)$. In the basic AFM (1), for a learner *i*, its coefficients consider the baseline of student proficiency (*α*), knowledge difficulty (*β*), and learning rates (*γ*) of previous exposures (*N*). The Performance Factor Model (PFM) (2) breaks down general exposure into success (*S*) and failure (*F*) for each knowledge unit (*α*) can be omitted in some versions (MacLellan et al., 2015), while the Instructional Factor Model (IFM) (3) goes beyond assessable steps and incorporates additional learning interactions such as direct tellings *T* from ITSs. Despite manifold variations, their core idea of cognitive modeling is to estimate learning rates for different knowledge units, resulting in potentially large numbers of parameters and domain-specific models. Efforts on model extensions often focus on expanding knowledge-specific indicators, such as forgetting models with slipping rates for each knowledge unit (MacLellan et al., 2015).

2.3 Connecting cognitive models to e-book context: adapted AFMs for behaviors

While the grounding systems and contexts differ, the two learning scenarios share similarities in data structure and problem space. Figure 3 illustrates this comparison, indicating the potential to apply AFMs in the L-K Modeling for OKLM construction. In AFM, the aim is to predict the probability of passing a step (L-S probability) using previous exposures to L-C and prescribed C-S. Similarly, for OKLM, the goal is to estimate knowledge proficiency (L-K) based on logs on L-R and prescribed K-R. This similarity allows for the reproducibility of AFM into the OKLM challenge.

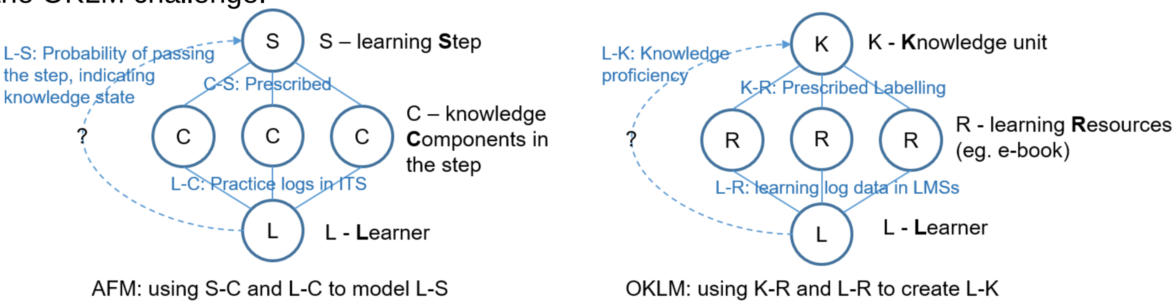

Figure 3. A comparison between learner modeling in OKLM and AFM.

However, differences exist between the two approaches. The original AFM emphasizes the effect of knowledge itself, while the OKLM context focuses on behavior adaptation. Therefore, in the adapted models discussed in this paper, we only employ behavior indicators from AFM models. We created formulas to illustrate the adapted AFMs tailored to the OKLM context, emphasizing learning rates of different learning behaviors beyond knowledge features.. Formula (4) derives from the basic AFM, (5) from PFM, and (6) from IFM respectively. Also, instead of “tell opportunities (*T*)” or other dialog acts in ITS in the original model, we use a more general term, such as engagement (*E*). For reference, we renamed each adapted model based on input behavior indicators (*S*, *F*, and *E*) in the formula and the subsequent description.

$$z_i = \sum_r (\gamma \times E(r,i)) \tag{4}$$

$$z_i = \sum_r (\mu \times S(r,i) + \rho \times F(r,i)) \tag{5}$$

$$z_i = \sum_r (\mu \times S(r,i) + \rho \times F(r,i) + \gamma \times E(r,i)) \tag{6}$$

Unlike ITS, there can be multiple learning resources in LA systems with limited reuse for each, making conducting AFM for individual resources less meaningful. Thus, to investigate a more general effect from a behavioral perspective, this study aims to further demonstrate the feasibility of estimation from behavior indicators by addressing the following research questions. We then conducted data analysis in a junior high math context to answer these questions. First, we examine the performance of the adapted model, followed by an exploration of its adaptability to different scenarios, spanning different knowledge units, learner levels, and learning stages.

- RQ1: Can we model knowledge proficiency based on behavior indicators with adapted AFM models?
- RQ2: Can the adapted AFM models fit across scenarios with different knowledge units and learning stages?

3. Research method

3.1 Learning context and data preparation

The study delves into junior high math, a scenario requiring cognitive skills and often learned through extensive practice in problem-solving. In Japan, a national project provides one tablet per student for ubiquitous and mobile learning (Sato & Uchiyama, 2023). In experimental schools where students were informed and consented to data collection, math practice problems are accessible via the BookRoll e-book system of LEAF (Ogata et al., 2024b), allowing students to engage with math problems during their learning. Each page of the book contains one math problem labeled with its corresponding knowledge units. The e-book learning unfolds in two contexts: during regular learning of new units and vacations for review purposes. Students can interactively operate the e-book with memos and hand-writings, and annotate their success or failure on each problem page as part of self-regulated learning practice.

The e-book system tracks success and failure in daily problem attempts, along with engagement indicators like annotations and memos. These indicators correspond to success (S) and failure (F) counts of steps in the IFM. Engagement (E) is counted based on whether the learner performs any operations on the page within a one-hour slot. All learning logs are anonymized when stored in the database with xAPI to deidentify learners. Additionally, to estimate the learning rate coefficients, actual grades of unit tests and post-vacation tests (where achieving 60% or more is considered passing) in the year 2023 were provided by the school teacher. These grades were deidentified and linked to system logs via system ID numbers assigned by the school.

The study employed the aforementioned math dataset with e-book learning logs as input and actual grades of tests as target output to construct the L-K proficiency model. A total of 49,423 records were employed after data cleaning. For the new knowledge learning, logs and grades from three geometry units were used from grade 3 students: “Similarity” (n = 64; 14,740 records), “Circle” (n = 31; 7,139 records), and “Pythagorean theorem” (n = 45; 12,269 records). For vacation review, logs and post-vacation grades from grade 1 students (n = 112; 15,275 records) were used, focusing on basic algebra, including rational numbers, algebraic expressions, and equations. Each data item records the total count of S, F, E, and the actual grade (scaled with the output probability) with a boolean flag indicating the pass obtained from the test of one student for one knowledge unit. Students with missing grades due to absence were excluded from the analysis.

3.2 Data analysis

For RQ1, we implemented adapted AFM, PFM, and IFM models, as depicted in Figure 3, using Python logistic regression tools (MacLellan et al., 2015). To assess the average model performance, we conducted 50 runs of 5-fold unstratified cross-validation for each of the three models using “Similarity” and “Circle” datasets. To assess the average model performance, we conducted 50 runs of 5-fold unstratified cross-validation for each of the three models using “Similarity” and “Circle” datasets. We compared the models in terms of confusion matrix-based **accuracy**, logit likelihood (**LL**), root mean square error (**RMSE**), and **correlation** between predicted probability and actual grade. As the number of parameters in our models is small, we did not consider the BIC. Regarding the learning rate coefficient, though some AFM variations restrict the coefficients of each knowledge to be positive under the premise that learners can learn even from errors (MacLellan et al., 2015), we do not set such bound limitation in our macro-level behavior.

For RQ2, we conducted two analyses to inspect the model adaptability across contexts with different spans. First, we evaluated the performance of the model trained in RQ1 using the “Pythagorean theorem” dataset, indicating a different knowledge unit and a relatively small context span. Subsequently, we tested the model using the summer vacation dataset, which represents a different learner level and learning stage and encompasses a larger context span.

As logit likelihood and RMSE are relative indicators, we examined accuracy, and correlations, and conducted paired sample tests between predicted probabilities and the actual grades of unit tests or post-vacation tests to ensure that the estimated knowledge proficiency reflects the actual knowledge.

4. Result

4.1 Performance under cross-validations

Table 1 summarizes the results of model construction from three approaches, presenting learning rate coefficients with 7 valid digits. In the IFM-based model (S-F-E), the learning rate of engagement (*E*) is tiny and negative, indicating a negligible contribution to final proficiency. Similarly, in the PFM-based model (S-F), the learning rate of failure practice exhibits a negative coefficient.

Table 1. *Averaged learning rate coefficients from cross-validations*

Adapted model	μ for S	ρ for F	γ for E
IFM	0.0223972	-0.0837946	-0.0054021
PFM	0.0104901	-0.0933654	
AFM			-0.0020906

Figure 4 shows the performance of three cognitive statistic models in the cross-validation of the training dataset. Each pair represents one run under a random data division from the same seed. While accuracy is similar across all models, S-F-E and S-F exhibit relatively higher rates. S-F demonstrates better performance in terms of RMSE and LL, whereas S-F-E shows poorer performance due to some extreme values. In terms of correlation to the actual grade, both S-F-E and S-F demonstrate positive relationships, though not statistically significant due to the small fold size in the cross-validation. The engagement-based model shows a negative correlation and performs poorly in this indicator to explain why was it less effective.

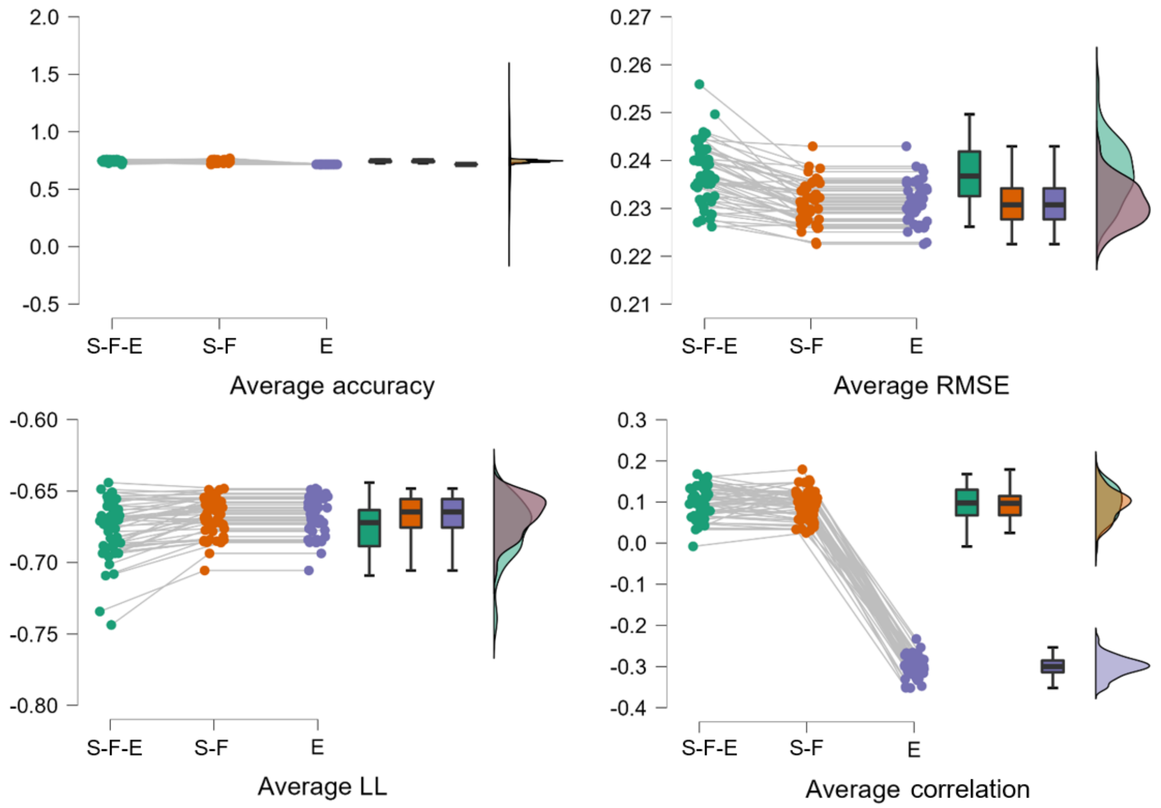

Figure 4. Performance of adapted AFM, PFM and IFM.

4.2 Adaptability to different learning contexts

As for the adaptability to different knowledge units, Figure 5 depicts the distribution of estimated proficiency compared to actual proficiency from the unit test. Each pair represents one learner in this case. Paired sample tests on two models confirm consistency ($p = .153 > .05$ for S-F-E and $p = .131 > .05$ for S-F) between the distributions at a confidence level of 95%. Despite a non-significant correlation coefficient, the models achieved accuracies of 0.89 for S-F-E and 0.91 for S-F, and the inaccurate predictions occur mainly in the positive true cell of the confusion matrix.

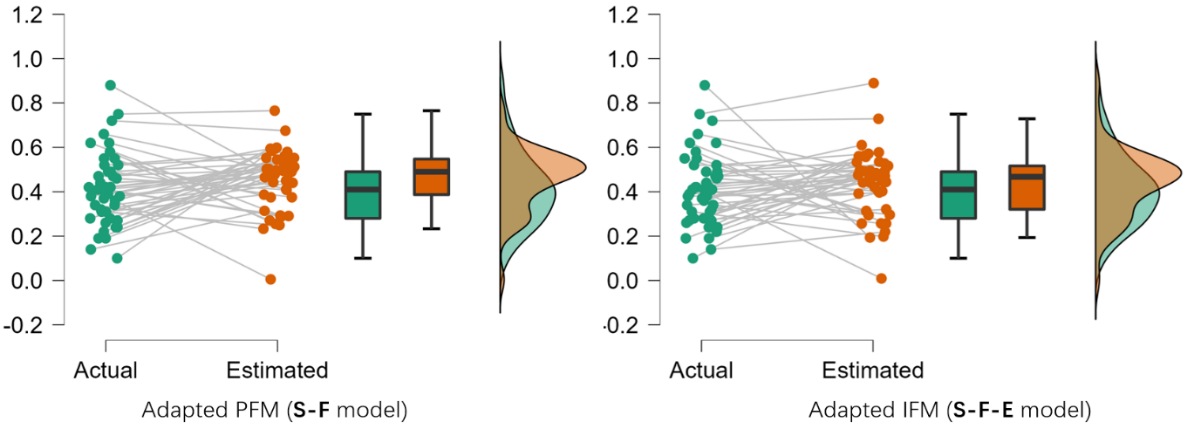

Figure 5. Distribution of estimated proficiency compared to actual proficiency from the unit test.

When applying the model to the dataset from a more diverse learning context with different learner levels and learning stages, the paired sample test revealed a significant difference between estimated and actual values ($p < .001$ for both S-F-E and S-F), suggesting a misalignment when the learning context varies with an accuracy of 0.18 for S-F-E and 0.17 for S-F. Figure 6 displays the original distribution of estimated proficiency in vacation practice compared to actual proficiency from the post-summer vacation test.

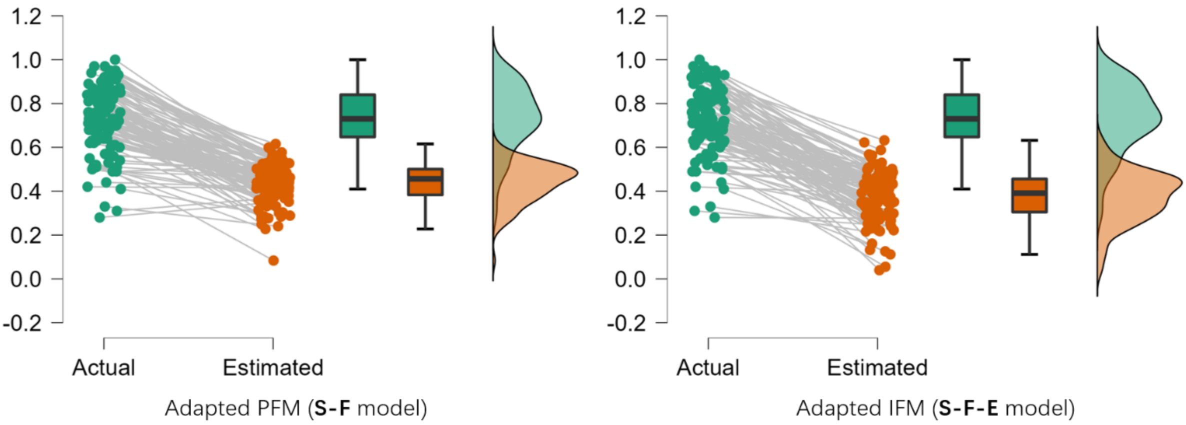

Figure 6. Original distribution of estimated proficiency compared to actual proficiency from post-summer vacation test.

Despite the significant difference, a notable positive correlation was observed between estimated and actual values ($r = .330$, $p < .01$ for S-F-E and $r = .412$, $p < .01$ for S-F). Given the disparity in average scores between unit tests and summer vacation practice, where the former represents initial exposure to the knowledge unit and the latter involves a review period, such discrepancies are explainable.

Evidently, the model's accuracy falls below acceptable levels. However, by adjusting the predicted probabilities to account for the difference in mean values between the two distributions, alignment across knowledge contexts was achieved (see Figure 7), with the model's accuracy improving to 0.74 for S-F-E and 0.82 for S-F.

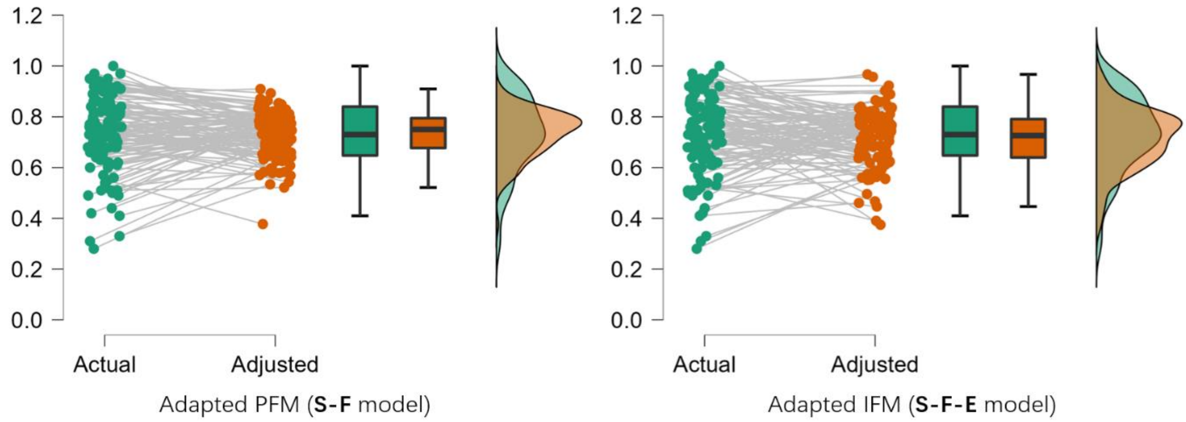

Figure 7. Distribution of adjusted proficiency compared to actual proficiency from post-summer vacation test.

5. Discussion

For RQ1, the results indicate that both adapted PFM (S-F) and IFM (S-F-E) can estimate knowledge proficiency at an acceptable level. However, AFM, relying solely on exposure counts, fails to reflect knowledge proficiency adequately. Unlike courses with low cognitive demand in higher education (Chen et al., 2021), mere engagement does not necessarily correlate with expected learning performance in math. The learning rate associated with pure engagement is small and negative, reflecting the phenomenon of wheel spinning commonly observed in math learning. It refers to a phenomenon where a student has spent considerable time practicing a skill, yet displays little or no progress towards mastery (Gong & Beck, 2015). These results underscore the imperative for nuanced selection of meaningful indicators from the abundant learning log data.

Additionally, The learning rate for success is the highest among the three indicators in the S-F-E and S-F models, while failure holds a negative learning rate. The finding is consistent with that in physics ITS (Chi et al., 2011), suggesting that subjects requiring higher knowledge might struggle to learn from errors. Moreover, it highlights the potential for LA studies on the wheel-spinning effect in cognitively demanding contexts, such as learning process graphs (LPG) based on knowledge models (Di Pierro et al, 2023).

Regarding RQ2, the results demonstrate the adaptability of the aforementioned models to broader scenarios. The initial analysis reveals that the model can successfully transition to other math knowledge units, affirming its versatility within a limited context. This suggests the possibility of utilizing the modeling approach to estimate learning proficiency in other new knowledge-learning contexts based on e-book math learning behaviors.

Although there is room for improvement in the positive true category in the confusion matrix due to the small sample of passed students, since our research goal is to estimate proficiency rather than solely classify pass or fail outcomes on the post-test, the model can be considered acceptable. Additionally, the analysis also indicates that the S-F model performed slightly better than the S-F-E model, which aligns with the cross-validation results in RQ1. Even minor differences like these could be crucial when selecting a model suitable for a specific application, ensuring effective practice and saving time for learners (Pelánek, 2017).

In the second analysis, it is acknowledged that model accuracy is affected by context differences. Training data for contexts where students encounter knowledge units for the first time exhibit diverse patterns compared to data from review stages during summer vacation. However, the model can adapt to new contexts with subtle adjustments. By incorporating an additional constant to account for the mean deviation of target grades between datasets, the adjusted model fits the new learning context effectively. This additional constant may represent the overall learner level depending on the learning stage, akin to the intercept of original AFM models (Chi et al., 2011). Moreover, the distribution of predicted proficiency aligns with actual performance, indicating that even original estimates of knowledge proficiency can apply to LA contexts concerned primarily with relative differences among learners, such as peer helper recommendation and group formation (Liang et al., 2024).

In comparison to traditional cognitive models that assign difficulty and learning rates to each knowledge concept, this study demonstrates the possibility of constructing models without considering knowledge differences. While the original models may yield higher accuracy and performance in specific contexts, our emphasis is on learner modeling using general behavior indicators for broader contexts, aligning with the principles of system interoperability advocated by the OKLM conception. Consequently, our adapted models help identify meaningful indicators from multiple behavior logs, determine the weight of each indicator to effectively model knowledge proficiency, and facilitate OKLM applications on a regular basis as outlined in Section 2.1.

The model in this study is still in its preliminary stage, using only behavior indicators related to success and failure from classical PFM models, which may oversimplify the assessment of knowledge proficiency. It is imperative to further disaggregate engagement (E) into more meaningful behavior aspects that could impact knowledge accumulation. This constitutes future work that the higher layers of cognitive processes can be explored. For instance, according to the definition of "tell" in the original IFM, logs on **peer help and feedback** could contribute to the knowledge model on the “evaluate” layer with certain weights (Vassileva et al., 2016). Besides, the potential of writing memos with **penstroke data** deserves further inspection to explore meaningful indicators for the “apply” and “analyze” layers involved in math learning (Yoshitake et al., 2020). Finally, the limited sample size may have resulted in some non-significant differences and the influence of extreme values in certain statistical analyses. Therefore, further data collection in consultation with experimental schools is necessary to address this limitation.

6. Conclusion

This study adapted cognitive statistical models based on logistic regression to model knowledge proficiency from everyday learning logs in digital learning systems. We emphasize the learning rates of general behavior indicators, departing from conventional models that focus on weights of specific domain knowledge. Through examination of authentic learning data from a junior high math e-book infrastructure, our study demonstrates the feasibility and adaptability of the models across diverse knowledge units and learning phases. Future work will explore more meaningful behavior indicators and realize inter-operable learner modeling in broader contexts.

Acknowledgements

This work was partially supported by CSTI SIP Program Grant Number JPJ012347 and NEDO JPNP20006.

References

Boticki, I., Akçapınar, G., & Ogata, H. (2019). E-book user modeling through learning analytics: The case of learner engagement and reading styles. *Interactive Learning Environments, 27*(5-6), 754–765.

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

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *International conference on intelligent tutoring systems* (pp. 164–175). Springer.

Chen, C.-H., Yang, S. J., Weng, J.-X., Ogata, H., & Su, C.-Y. (2021). Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers. *Australasian Journal of Educational Technology, 37*(4), 130–144.

Chi, M., Koedinger, K., Gordon, G., Jordan, P., & VanLehn, K. (2011). Instructional factors analysis: A cognitive model for multiple instructional interventions. In *4th International Conference on Educational Data Mining* (pp. 61–70).

Chounta, I.-A., Albacete, P., Jordan, P., Katz, S., & McLaren, B. M. (2017). The "grey area": A computational approach to model the zone of proximal development. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, Proceedings 12* (pp. 3–16). Springer.

Di Pierro, D., Ferilli, S., & Redavid, D. (2023). LPG-based knowledge graphs: A survey, a proposal and current trends. *Information, 14*(3), 154.

Draney, K. L., Pirolli, P., & Wilson, M. (2012). A measurement model for a complex cognitive skill. In P. Nichols (Ed.), *Cognitively diagnostic assessment* (pp. 103–125). Routledge.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*(2), 213–232.

Gong, Y., & Beck, J. E. (2015). Towards detecting wheel-spinning: Future failure in mastery learning. In *Proceedings of the second (2015) ACM conference on learning@scale* (pp. 67–74).

Khadir, A. C., Aliane, H., & Guessoum, A. (2021). Ontology learning: Grand tour and challenges. *Computer Science Review, 39*, 100339.

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

Liang, C., Majumdar, R., Horikoshi, I., Ogata, H. (2024). Data-Driven Support Infrastructure for Iterative Team-Based Learning. *IEEE Access*, *12*, 65967-65980.

MacLellan, C. J., Liu, R., & Koedinger, K. R. (2015). Accounting for slipping and other false negatives in logistic models of student learning. In *Proceedings of the 8th International Conference on Educational Data Mining*. ERIC.

Ogata, H., Flanagan, B., Takami, K., Dai, Y., Nakamoto, R., & Takii, K. (2024a). EXAIT: Educational explainable artificial intelligent tools for personalized learning. *Research and Practice in Technology Enhanced Learning, 19*.

Ogata, H., Majumdar, R., Flanagan, B., & Kuromiya, H. (2024b). 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.

Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction, 27*, 313–350.

Sato, H., & Uchiyama, E. (2023). Digital transformation in primary and secondary education in japan. In *Emerging Trends and Historical Perspectives Surrounding Digital Transformation in Education: Achieving Open and Blended Learning Environments* (pp. 177–199). IGI Global.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai. *International Journal of Human-Computer Studies, 146*, 102551.

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 Learning Analytics and Knowledge Conference* (pp. 191–193).

Vassileva, J., McCalla, G. I., & Greer, J. E. (2016). From small seeds grow fruitful trees: How the phelps peer help system stimulated a diverse and innovative research agenda over 15 years. *International Journal of Artificial Intelligence in Education, 26*, 431–447.

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. Harvard University Press.

Wang, Y., Kai, S., & Baker, R. S. (2020). Early detection of wheel-spinning in assistments. In *Artificial* Intelligence *in Education: 21st International Conference, AIED 2020, Proceedings, Part I 21* (pp. 574–585).

Yoshitake, D., Flanagan, B., & Ogata, H. (2020). Supporting Group Learning Using Pen Stroke Data Analytics. *In 28th International Conference on Computers in Education Conference Proceedings* (Vol. 1, pp. 634-639).