Multi-Feature Knowledge Tracing with Time-Series Interpretability for Mathematics Teacher Micro-Credentials

Hai Li
University of Florida
li.ha@ufl.edu

Chenglu Li
University of Utah
chenglu.li@utah.edu

ABSTRACT

Teacher professional development (PD) is critical for instructional improvement, and online micro-credentials provide flexible, work-embedded learning opportunities. Knowledge tracing (KT) can support personalization by modeling knowledge state evolution, but systematic application to adult professional learners remains limited, particularly when interaction logs are multi-dimensional and models must balance predictive performance with interpretability under small-sample constraints. Focusing on mathematics teacher micro-credentials, we propose a multi-feature deep KT framework with built-in temporal interpretability. We organize heterogeneous log-derived features into a progressive six-layer hierarchy and apply mutual information-based selection, which is well-suited for small-sample, feature-rich environments. Our residual-injection Time Interpretation Module (TIM) augments LSTM-based Deep Knowledge Tracing with lightweight temporal encoding and causal self-attention, providing intrinsic interpretability through temporal feature importance analysis. Evaluated on a dataset comprising 154 teachers and 22,988 interactions, results demonstrate that compact core-feature configurations achieve strong predictive performance, while indiscriminate feature addition degrades accuracy due to dimensionality effects. TIM improves both predictive performance and training stability while revealing interpretable temporal dynamics distinctive to adult professional learners. The framework offers systematic guidance for KT-based support in resource-constrained professional learning contexts.

Keywords

Knowledge Tracing, Teacher Professional Development, Time-Series Interpretability, Multi-Feature Integration

1. INTRODUCTION

Teacher professional development (PD) is widely viewed as a high-leverage route to improving instruction and student outcomes [29, 6]. Contemporary PD increasingly emphasizes personalization, work-embedded learning, and competency-oriented certification through *micro-credentials* that enable modular, flexible asynchronous participation [7, 22]. A central goal is strengthening *Pedagogical Content Knowledge (PCK)* in authentic instructional contexts [3].

Knowledge Tracing (KT) offers data-driven modeling of knowledge state evolution and can underpin adaptive support systems [5, 25]. The field has progressed from Bayesian Knowledge Tracing (BKT) and Performance Factors Analysis (PFA) [24] to LSTM-based Deep Knowledge Tracing (DKT) [25] and attention-augmented architectures [30, 23, 21]. Recent advances have further enhanced KT through multi-time series feature integration [13] and option-level tracing for mathematics assessment [15]. LSTM-based models combined with attention mechanisms represent well-established approaches for capturing temporal dependencies in learner sequences [1]. However, systematic application to adult professional learners remains limited, particularly in asynchronous micro-credential settings where learning rhythms, communicative patterns, and constraints differ substantially from traditional educational environments [19].

Teachers, as adult learners, exhibit autonomy, experience-based reasoning, and problem-centered learning consistent with *andragogy* [9]. Their professional learning involves complex instructional judgments and deep PCK integration [20]. Using data from a mathematics teacher micro-credential program, we observe distinctive properties: highly concentrated sessions, notably high item difficulty (overall accuracy 33.37%), and substantial variation in skill mastery across 12 teaching dimensions. These characteristics suggest that systematic feature engineering and temporal modeling are particularly important in this context.

Methodologically, teacher PD platforms generate multi-dimensional logs encompassing behavioral, temporal, and performance signals, yet feature processing in existing KT work is often ad hoc [11]. Small-sample, feature-rich settings amplify overfitting risks and dimensionality issues. Additionally, many deep KT models lack intrinsic interpretability for temporal dynamics [1], limiting their utility in professional learning contexts where stakeholders need to understand prediction mechanisms.

We address two research questions that systematically tackle these challenges:

Hai Li, and Chenglu Li. Multi-Feature Knowledge Tracing with Time-Series Interpretability for Mathematics Teacher Micro-Credentials. In Anthony Botelho, Maria Mercedes T. Rodrigo, Adish Singla, Hiroaki Ogata, Hyojeong So, and Young Hoan Cho (eds.) Proceedings of the 19th International Conference on Educational Data Mining, Seoul, Republic of Korea, June, 2026, pp. 715–720. International Educational Data Mining Society (2026).

https://doi.org/10.5281/zenodo.21039794

RQ1: How do different log-derived features, organized via a progressive hierarchy and selected through mutual information, affect KT predictive performance in a small-sample professional learning setting?

RQ2: How can a lightweight, residual-injected temporal module enhance LSTM-based DKT performance while providing interpretable temporal dynamics?

2. RELATED WORK

2.1 Knowledge Tracing and Temporal Modeling

KT has evolved from BKT's Hidden Markov Models [5] and PFA's logistic regression with engineered predictors [24] to DKT's recurrent sequence modeling [25]. LSTM-based architectures combined with attention mechanisms are well-established for capturing temporal dynamics in learner modeling [30, 23, 21, 1]. Models such as Dynamic Key-Value Memory Networks (DKVMN) have introduced structured memory mechanisms that provide specific levels of interpretability [21]. Temporal modeling and forgetting effects remain active areas of development [4].

Despite these advances, most KT evidence derives from student learning contexts. Teachers as adult learners follow andragogy principles including self-direction and experience integration, implying KT models must accommodate higher-order cognitive processes distinct from student content learning. Micro-credentials further amplify these differences through competency-oriented certification and asynchronous pacing [22].

2.2 Multi-Feature Integration and Interpretability

As learning platforms expand, KT research increasingly integrates process data such as response time, hints, and navigation paths to improve prediction [1, 17, 2]. Parallel developments in mathematics education have demonstrated the value of multimodal features for automated feedback [12] and collaborative filtering approaches for automatic scoring [16]. However, several methodological challenges persist: feature construction for heterogeneous logs remains largely unstructured, with limited systematic use of nonlinear dependency metrics such as mutual information [10]; temporal features are often incorporated via simple concatenation, which struggles to capture complex sequential effects [4]; and model interpretability typically relies on post-hoc tools such as SHAP or LIME [18, 27], making it difficult to isolate the role of specific temporal features [28].

3. METHOD

We model teacher learning as sequence $\mathcal{S} = \{(s_t, q_t, r_t)\}_{t=1}^{T}$, where s_t denotes skill ID ($K = 12$), q_t question ID ($Q = 127$), and $r_t \in \{0, 1\}$ binary response. The prediction goal is $p_t = P(r_t = 1 \mid \mathcal{S}_{<t}, \boldsymbol{f}_t)$.

3.1 Platform Context and Dataset Construction

The data originates from an asynchronous mathematics teacher micro-credential platform deployed in the southeastern United States (September 2024–February 2025). The platform supports practicing educators working toward competency-based certification through modular courses. Assessment items consist of complex pedagogical judgment tasks where teachers analyze authentic classroom scenarios and identify specific instructional strategies such as revoicing, fostering reasoning, and leveraging student errors. These higher-order tasks target PCK integration, which accounts for the dataset's notably low overall accuracy and substantial cross-skill variation.

We align four heterogeneous data sources at teacher and course levels: assessment response records, content learning progress, platform interaction logs, and course metadata. The learning environment follows a standard dashboard-content-assessment workflow where teachers access materials, attempt items, and receive feedback within an integrated web platform. All data were de-identified prior to analysis under institutional data agreements and ethics review protocols. Preprocessing follows three rigorous principles: features with $> 70\%$ missingness are dropped to limit imputation bias; response-time outliers (< 1s or > 3600s) are removed; and strict temporal ordering is enforced via cumulative `time_since_last_attempt` to prevent future leakage.

Dataset Summary: The final dataset comprises **22,988 valid interaction records from 154 teachers**, covering **127 items across 12 skills**, with overall accuracy 33.37% and average sequence length 149.3 ± 30.2 items per teacher. Temporal characteristics reflect concentrated adult learning behavior: **99.84% of consecutive responses occur within sessions** (< 6 minutes apart), while long intervals (> 1 week) account for only 0.14% of transitions.

3.2 Training Protocol and Statistical Rigor

Given the relatively small teacher cohort (**154 teachers**) paired with high interaction volume (**22,988 records**), teacher-level generalization is the primary evaluation concern. We apply teacher-level stratified splits with complete isolation: **Training (107 teachers, 16,065 records, 69.9%), Validation (23 teachers, 3,538 records, 15.4%), Test (24 teachers, 3,385 records, 14.7%)**. This ensures evaluation reflects true generalization to unseen teachers.

Optimization uses Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$, lr=0.001) with class weights (pos_weight=1.99) to address class imbalance, label smoothing ($\epsilon = 0.02$), gradient clipping (max_norm=0.8), and early stopping (patience=5). To ensure robust statistical reporting, **each configuration runs with 5 independent random seeds (42–46)**, with results reported as empirical mean $\pm$ standard deviation. We use variance reduction across seeds as the primary stability indicator, which provides reliable evidence of model improvements under deep learning initialization variability.

3.3 Progressive Feature Engineering (RQ1)

We organize features into a six-layer hierarchy from core to contextual, enabling controlled evaluation of each feature group's contribution under small-sample constraints (Table 1). This design systematically addresses the practical question of which feature subsets offer optimal performance-cost tradeoffs for deployment.

Mutual information $I(f; r) = \sum_{f,r} p(f, r) \log \frac{p(f,r)}{p(f)p(r)}$ is com-

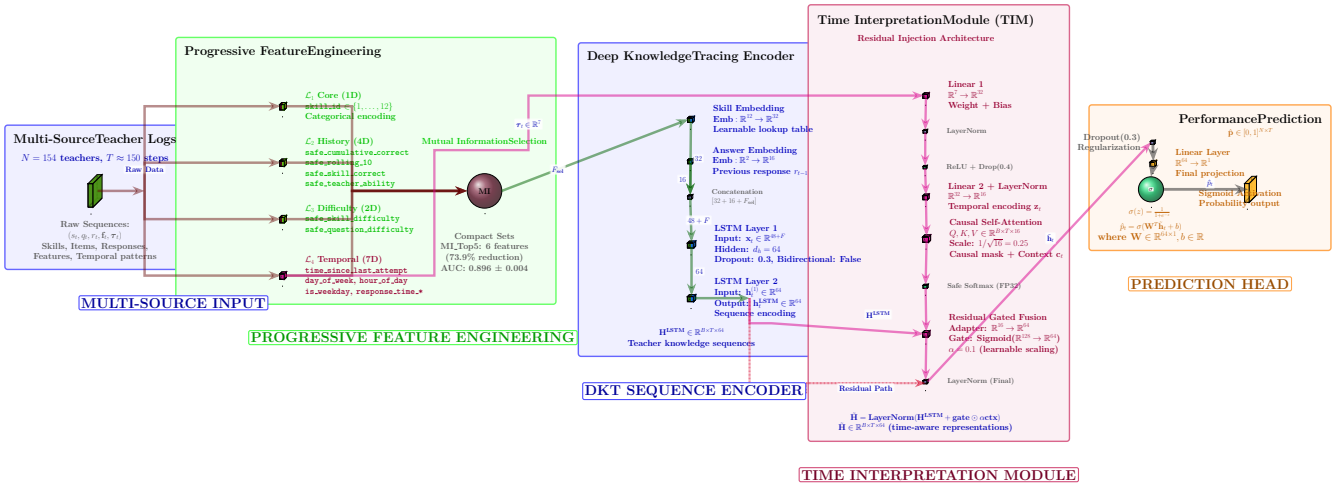

Figure 1: Overall architecture of the multi-feature DKT framework with residual-injection Time Interpretation Module (TIM).

Table 1: Progressive Feature Hierarchy System

Layer	Category	Variables & Description	No.
$\mathcal{L}_1$ *Core (L1_Baseline)*			
	Skill	`skill_id`: 12 categories, 32-dim embeddings	1
$\mathcal{L}_2$ *Performance History*			
	History	`cumulative_accuracy`, `rolling_accuracy_10`, `skill_accuracy`, `teacher_ability`	4
$\mathcal{L}_3$ *Difficulty Parameters*			
	Difficulty	`item_difficulty`, `skill_difficulty`	2
$\mathcal{L}_4$ *Temporal Features*			
	Temporal	`time_since_last_attempt`, `hour_of_day`, `day_of_week`, `is_weekday`, `response_time_seconds`, `response_time_log`, `response_time_zscore`	7
$\mathcal{L}_5$ *Behavioral Features*			
	Behavioral	`total_learning_time`, `total_visits`, `avg_time_per_visit`, `questions_per_hour`, `total_attempts_by_user`	5
MI Selection (MI_TopK)			
	Selection	Top-K by mutual information (K=5,10,15)	6,11,16

Full configuration: L1(1)+L2(4)+L3(2)+L4(7)+L5(5)=19 total features.

puted via k-nearest-neighbor estimation [10], ranking 18 non-core features to construct MI_Top5/10/15 configurations. This nonlinear dependency metric is particularly effective for capturing complex feature-response relationships in small-sample, feature-rich settings where linear correlation may underestimate feature relevance.

3.4 LSTM-based DKT with Time Interpretation Module (RQ2)

The base DKT concatenates skill embedding $\boldsymbol{e}_s(s_t) \in \mathbb{R}^{32}$, answer embedding $\boldsymbol{e}_a(r_{t-1}) \in \mathbb{R}^{16}$, and auxiliary features $\boldsymbol{f}_t$ into input $\boldsymbol{x}_t$. A 2-layer LSTM (hidden size 64, dropout 0.3) encodes $\boldsymbol{h}_t = \text{LSTM}(\boldsymbol{x}_t, \boldsymbol{h}_{t-1})$, with prediction $p_t = \sigma(\boldsymbol{w}^\top \boldsymbol{h}_t + b)$.

Table 2: Model Performance Analysis (mean ± std, 5 runs, sorted by AUC)

Configuration	Feat.	AUC	F1	Acc.
BKT (Baseline)	1	0.445 ± 0.000	0.122 ± 0.000	66.5%
PFA (Full)	19	0.769 ± 0.000	0.508 ± 0.000	73.4%
DKT (Full)	19	0.777 ± 0.008	0.610 ± 0.004	67.5%
DKT (MI_Top15)	16	0.894 ± 0.011	0.749 ± 0.016	83.0%
DKT (MI_Top5)	6	0.894 ± 0.006	0.750 ± 0.012	82.9%
DKT (L1_Baseline)	1	$\mathbf{0.896 \pm 0.004}$	$\mathbf{0.752 \pm 0.003}$	**83.0%**

Time Interpretation Module (TIM) extends the base DKT with temporal interpretability via residual injection, comprising four components: (1) *Temporal encoder*: a 2-layer MLP with LayerNorm maps raw temporal features $\boldsymbol{\tau}_t \in \mathbb{R}^7$ to compact encoding $\boldsymbol{z}_t \in \mathbb{R}^{16}$; (2) *Causal self-attention* [28] with scaling factor $\sqrt{16}$ captures temporal dependencies across sequence steps; (3) *Gradient-based feature importance estimation* provides step-level interpretability of temporal feature contributions; (4) *Residual gated fusion*: $\tilde{\boldsymbol{c}}_t = \alpha \cdot \boldsymbol{c}_t$, $\boldsymbol{g}_t = \sigma(\boldsymbol{W}_g[\boldsymbol{h}_t; \tilde{\boldsymbol{c}}_t] + \boldsymbol{b}_g)$, $\boldsymbol{h}'_t = \text{LayerNorm}(\boldsymbol{h}_t + \boldsymbol{g}_t \odot \tilde{\boldsymbol{c}}_t)$. The gating mechanism adaptively controls temporal signal contribution, providing stable integration while preserving baseline DKT representational capacity.

4. RESULTS

4.1 RQ1: Progressive Feature Integration and Performance

Mutual information ranking reveals the strongest individual associations with correctness: `skill_accuracy` (MI=0.064), `item_difficulty` (0.030), and `cumulative_accuracy` (0.028). Temporal features show lower individual MI values, indicating their contribution depends on sequential relational modeling.

Three systematic patterns emerge from Table 2. First, *core-feature sufficiency*: L1_Baseline achieves optimal AUC

Table 3: TIM Ablation Study (based on DKT MI_Top15, 5 runs)

Configuration	AUC	F1	Std Change
TIM w/o attention	0.821 ± 0.021	0.657 ± 0.027	+91%
DKT (baseline)	0.894 ± 0.011	0.749 ± 0.016	ref
TIM w/o importance	0.921 ± 0.009	0.801 ± 0.022	−18%
TIM (complete)	$\mathbf{0.925 \pm 0.002}$	$\mathbf{0.805 \pm 0.005}$	**−82%**

(0.896), reflecting the strong predictive signal embedded in the inherent difficulty hierarchy across the 12 PCK skills, which LSTM captures efficiently from skill identity alone. Second, *dimensionality effects*: DKT (Full) with all 19 features drops to AUC=0.777, comparable to PFA, demonstrating that adding noisy features without selection introduces overfitting in this small-sample setting. Third, *selection effectiveness*: MI_Top5 (6 features) nearly matches the skill-only ceiling while outperforming the full feature set by 15.1% in AUC, confirming that principled mutual-information selection recovers predictive value lost to dimensionality.

4.2 RQ2: TIM Architecture and Temporal Interpretability

Since L1_Baseline contains only skill identity and cannot evaluate temporal modeling, we use DKT (MI_Top15) as the ablation baseline, as it includes all 7 temporal features while maintaining competitive performance.

Three significant findings emerge from Table 3. First, *robust temporal integration*: Complete TIM improves AUC to 0.925 (+3.1 points over baseline) while reducing standard deviation by 82%, surpassing the skill-only ceiling (0.896) and confirming that temporal features carry genuine predictive value when modeled via gated residual fusion. Second, *attention criticality*: Removing causal self-attention causes severe degradation (AUC=0.821, std +91%), demonstrating that temporal features require relational modeling across sequence steps to contribute reliably. Third, *importance estimation*: Removing this component yields modest performance decline, confirming it primarily serves interpretability with limited direct predictive impact.

Temporal feature importance: TIM assigns highest importance to time-arrangement features: `day_of_week` (0.193 ± 0.103), `hour_of_day` (0.187 ± 0.082), and `is_weekday` (0.171 ± 0.078), while spacing intervals such as `time_since_last_attempt` (0.156 ± 0.059) receive comparatively lower weights. Attention distributions show stable 77% weight assigned to distant history (≥ 5 steps, std=0.000), indicating consistent reliance on accumulated experience over recent fluctuations, consistent with schema-based reasoning in adult learners.

5. DISCUSSION

5.1 Systematic Synthesis of Findings

The two research questions together reveal a coherent understanding of teacher KT in micro-credential settings. For RQ1, the strong performance of compact feature sets reflects the structured difficulty hierarchy inherent in PCK assessments: skill identity alone encodes substantial predictive signal, and adding heterogeneous features without principled selection introduces more noise than value in small-sample conditions. The mutual information selection approach effectively recovers this signal by retaining performance-relevant features while discarding redundant ones.

For RQ2, TIM's improvements demonstrate that temporal features carry genuine predictive value, but only when modeled with appropriate sequential structure. The gated residual design provides a stable integration pathway that incorporates temporal context when informative and attenuates it when weak, enabling reliable performance gains across random seeds. The temporal importance analysis reveals that scheduling context (when teachers engage) is more informative than spacing intervals (how long between sessions), suggesting that adult learners' within-day engagement patterns carry meaningful predictive signal in concentrated learning settings.

5.2 Practical Implications for Professional Development

These findings offer systematic guidance for deploying KT in professional learning contexts. The MI_Top5 configuration achieves near-optimal performance with only 6 features, providing a favorable performance-cost tradeoff for resource-constrained platforms. TIM's residual gating provides robustness to variable temporal signal quality across teachers, making it suitable for heterogeneous adult learner populations. Platform designers may prioritize logging scheduling context (time of day, day of week) as high-value signals for future KT integration, while recognizing that skill-based features provide the strongest foundational predictive capacity. To illustrate pedagogical applications, consider a teacher whose TIM-predicted mastery probability in "leveraging student errors" remains below 0.4 despite progress in other PCK dimensions. The model's temporal importance analysis reveals this teacher typically attempts related items late at night with inconsistent engagement patterns. A PD dashboard could surface this insight by highlighting the lagging skill, visualizing predicted trajectories, and recommending targeted micro-lessons scheduled during time windows when the model estimates higher success probability for this individual teacher. Additionally, content-level adaptations such as readability optimization of mathematical materials can further enhance learner engagement [14].

5.3 Limitations and Future Directions

This study has several constraints that define important directions for future work. First, the dataset derives from a single mathematics micro-credential program, and generalizability to other subject domains or PD modalities requires further validation. Second, our analyses rely on structured interaction logs and do not incorporate richer modalities such as open-ended responses or classroom artifacts, which could deepen pedagogical insight.

Third, the current framework treats the 12 PCK dimensions as distinct tracks, whereas in complex professional development, these instructional strategies (e.g., fostering reasoning vs. leveraging student errors) likely interact dynamically. Exploring how mastery of one specific pedagogical skill sequentially influences the acquisition of another—perhaps through network analysis techniques—could provide deeper understanding of how these competencies co-evolve. Fourth,

while TIM provides model-level temporal interpretability, validation with practicing teachers is needed to assess how these model-derived explanations align with teachers' subjective experiences.

Finally, although we establish clear baselines against BKT, PFA, and standard DKT, comparison with more contemporary attention-based models such as SAKT or AKT would make the performance claims significantly more compelling. The current small-sample constraint (154 teachers) poses practical challenges for training data-intensive transformer architectures; extending these models to accommodate multi-dimensional temporal features [26] with robust optimization strategies [8] under such constraints represents a promising direction for future work.

6. CONCLUSION

This study presents a systematic multi-feature KT framework with intrinsic temporal interpretability for mathematics teacher micro-credentials. A six-layer progressive feature hierarchy with mutual information selection addresses the small-sample, feature-rich challenge, demonstrating that compact MI-selected configurations outperform full feature sets by recovering predictive value lost to dimensionality. A residual-injection Time Interpretation Module extends LSTM-based DKT with stable temporal fusion and gradient-based importance analysis, improving AUC by 3.1 points while reducing training variance by 82%. Temporal analysis reveals that scheduling context and accumulated experience are primary predictive signals in this concentrated adult learning setting. The framework provides a reusable, principled approach to KT-based support for teacher professional development that balances predictive performance with interpretability requirements.

7. ACKNOWLEDGMENTS

The research reported here was supported by AIMS EduData through the Gates Foundation, the Lastinger Center for Learning at the University of Florida, and the One-U Responsible AI Faculty Fellowship at the University of Utah. The views expressed are those of the authors and do not necessarily reflect those of the funders or the University of Utah.

8. REFERENCES

[1] G. Abdelrahman, Q. Wang, and B. Nunes. Knowledge tracing: A survey. *ACM Computing Surveys*, 55(11):1–37, 2023.

[2] S. Blömeke, R. T. Houang, and U. Suhl. Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models. In *International perspectives on teacher knowledge, beliefs and opportunities to learn: TEDS-M results*, pages 483–501. Springer, 2014.

[3] J. Carlson, K. R. Daehler, A. C. Alonzo, E. Barendsen, A. Berry, A. Borowski, J. Carpendale, K. Kam Ho Chan, R. Cooper, P. Friedrichsen, et al. The refined consensus model of pedagogical content knowledge in science education. In *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science*, pages 77–94. Springer, 2019.

[4] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, 132(3):354, 2006.

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

[6] L. Darling-Hammond, M. E. Hyler, and M. Gardner. Effective teacher professional development. *Learning policy institute*, 2017.

[7] L. M. Desimone. Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher*, 38(3):181–199, 2009.

[8] D. P. Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[9] M. S. Knowles and Associates. *Andragogy in Action: Applying Modern Principles of Adult Learning*. Jossey-Bass, San Francisco, CA, 1984.

[10] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 69(6):066138, 2004.

[11] N. A. Levin et al. Process mining combined with expert feature engineering to predict efficient use of time on high-stakes assessments. *Journal of Educational Data Mining*, 13(2):1–15, 2021.

[12] H. Li, C. Li, W. Xing, S. Baral, and N. Heffernan. Automated feedback for student math responses based on multi-modality and fine-tuning. In *Proceedings of the 14th learning analytics and knowledge conference*, pages 763–770, 2024.

[13] H. Li and W. Xing. Enhanced knowledge tracing: Leveraging multi-time series features from interaction process via an attention-based framework. In *International Conference on Artificial Intelligence in Education*, pages 250–258. Springer, 2025.

[14] H. Li, W. Xing, C. Li, W. Zhu, and H. Oh. Are simpler math stories better? automatic readability assessment of gai-generated multimodal mathematical stories validated by engagement. *British Journal of Educational Technology*, 56(3):1092–1117, 2025.

[15] H. Li, W. Xing, C. Li, W. Zhu, and S. Woodhead. Integrating option tracing into knowledge tracing: enhancing learning analytics for mathematics multiple-choice questions. *Journal of Learning Analytics*, 12(1):322–337, 2025.

[16] H. Li, W. Xing, W. Zhu, C. Li, B. Lyu, Z. Liu, and N. Heffernan. Leveraging multi-modality and collaborative filtering for supporting automatic scoring in mathematics education. In *International Conference on Artificial Intelligence in Education*, pages 313–320. Springer, 2025.

[17] Y. Lu, L. Tong, and Y. Cheng. Advanced knowledge tracing: Incorporating process data and curricula information via an attention-based framework for accuracy and interpretability. *Journal of Educational Data Mining*, 16(2):58–84, 2024.

[18] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[19] B. Lyu, C. Li, H. Li, and W. Xing. Exploring students' participation in online mathematical discussions using educational big data: A communicative ecology perspective. *Educational Technology & Society*, 29(2):171–190, 2026.

[20] B. L. Moore-Adams, W. M. Jones, and J. Cohen. Learning to teach online: A systematic review of the literature on k-12 teacher preparation for teaching online. *Distance education*, 37(3):333–348, 2016.

[21] H. Nakagawa, Y. Iwasawa, and Y. Matsuo. Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *IEEE/WIC/aCM international conference on web intelligence*, pages 156–163, 2019.

[22] National Education Association. Micro-credential guidance for certified educators. Technical report, National Education Association, 2021.

[23] S. Pandey and G. Karypis. A self-attentive model for knowledge tracing. *arXiv preprint arXiv:1907.06837*, 2019.

[24] P. I. Pavlik Jr, H. Cen, and K. R. Koedinger. Performance factors analysis–a new alternative to knowledge tracing. *Online submission*, 2009.

[25] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. Deep knowledge tracing. *Advances in neural information processing systems*, 28, 2015.

[26] Y. Qiu, Y. Qi, H. Lu, Z. A. Pardos, and N. T. Heffernan. Does time matter? modeling the effect of time with bayesian knowledge tracing. In *Edm*, pages 139–148, 2011.

[27] M. T. Ribeiro, S. Singh, and C. Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[29] K. S. Yoon, T. Duncan, S. W.-Y. Lee, B. Scarloss, and K. L. Shapley. Reviewing the evidence on how teacher professional development affects student achievement. issues & answers. rel 2007-no. 033. *Regional Educational Laboratory Southwest (NJ1)*, 2007.

[30] J. Zhang, X. Shi, I. King, and D.-Y. Yeung. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, pages 765–774, 2017.