IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume 11,Issue 4, 2023

A Survey on Optimal Learning Models to Monitor Student Progress in Knowledge Tracing

D.Divya Sai Chand, M.Tech(AI & DS)

ABSTRACT:

Understudies' ability to study in web-based learning frameworks has improved as a result of Knowledge tracing (KT), which refers to following understudies' shifting information state as they learn. Due to its fundamental significance in education, KT has recently attracted considerable examination attention. Nevertheless, the bulk of contemporary KT tactics aim for high accuracy in understudy execution expectations while ignoring the consistency between understudies' shifting information states and their learning styles. In this article, we explore an alternative worldview for the KT job and offer a creative model called Learning process consistent Knowledge Tracing (LPKT) and LPKT-S that checks learners' knowledge levels by plainly demonstrating their preferred method of learning. In specifically, the fundamental learning cell is initially formalised as the tuple practise answer time reply. Then, using the difference between the present and past learning cells, their duration, and the related information state of the students, we carefully measure the learning gain as well as its diversity. In order to determine the amount of information that students can absorb, we also design a learning door. In addition, we design an ignoring door to show how understudies' information deteriorates over time depending on their prior information condition, current learning gains, and time span. Broad testing findings on an open dataset demonstrate how LPKT could obtain more accurate information state as experience grows. Furthermore, LPKT also outperforms modern KT techniques in terms of expected understudy performance. Our work illustrates a potential future test bearing for KT that has a high level of precision and interpretability.

KEYWORDS: knowledge tracing, learning process, learning gain, forgetting effect.

I.INTRODUCTION:

Online education [1] has expanded rapidly in recent years and is essential to enhancing education [2,3]. Knowledge tracing (KT) [4], a new area of research in online learning, uses machine learning sequence models that can monitor students' changing knowledge states using educationally relevant data. By completing several activities in the online learning system, students can grasp their knowledge.

M.Tech(AI & DS)
Chaitanya Bharathi Institute of Technology, Hyderabad, India

The KT task, in turn, formalizes our capacity to infer students' knowledge levels [5,6,7,8] and forecast their future performance based on their learning sequences. According to the students' prior learning sequence, which includes exercises and replies, the KT task [4,9] explicitly tries to measure the knowledge states of students at various time points, which can be used to predict future performance. Students and teachers will be able to focus more on topics that are well-mastered rather of spending time on those that are not after they have a better understanding of their knowledge states [10], which will take some time. KT can benefit both teaching and learning at the same time.

II.LITERATURE SURVEY:

Deep learning, logistic models, and probabilistic models are the foundations of the majority of the KT models that are now accessible. The popular probabilistic model for KT, BKT [4] is one instance of a specific use of the Hidden Markov Model (HMM). In a vast family of logistic function-based models, such as Performance Factor Analysis (PFA) [21], a logistic function was employed to calculate the likelihood of a knowledge state. DKT was the first to incorporate deep learning into KT [12]. The learning sequence was used as the input for an RNN or its variant, the LSTM, and hidden states were used to represent the student knowledge states in this method.

The first memory-augmented neural networks, known as Dynamic Key-Value Memory Networks (DKVMN), were presented to KT [22]. It established a key—a static matrix—and a value—a dynamic matrix—for storing and enhancing knowledge mastery for latent knowledge ideas [22]. To improve the task's effectiveness, EKT included text components.

Convolutional windows were utilised in convolutional knowledge tracing (CKT) to characterise the different continuous learning interactions' individual learning rates for each student [23]. The self-attention mechanism was used in the self-aware model for knowledge tracing to replicate SAKT and the long-term interdependence between learning interactions [24]. Pandey and Srivastava created a relation-aware self-attention layer that considers the context [25]. The context-aware attentive knowledge tracing (AKT) model was developed by Ghosh and colleagues [26] by fusing cognitive and psychometric models with contextualized representations of both exercises and knowledge acquisition. This model included an attention mechanism. Shen and colleagues looked on how challenging questions affected students' learning in KT. It can be used by readers of Liu & Co. and Schmucker et al. and to conduct a detailed analysis of current KT developments [28].

2.5 Different Models of Knowledge Tracing

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 11,Issue 4, 2023

Model	Description
Bayesian Knowledge Tracing (BKT)	A traditional knowledge tracing model that models the probability of a student mastering a skill based on their past performance.
Factor Analysis Models (FAM)	Another traditional knowledge tracing model that models the correlation between skills and the probability of a student mastering them.
Deep Knowledge Tracing (DKT)	A deep learning model that uses a recurrent neural network to model the probability of a student mastering a skill based on their past performance.
Self-Attentive Knowledge Tracing (SAKT)	A deep learning model that uses self-attention to model the probability of a student mastering a skill based on their past performance.
Key-Value Memory Network (KVMN)	A deep learning model that uses a memory network to model the probability of a student mastering a skill based on their past performance.

Table.1. Different Models of Knowledge Tracing

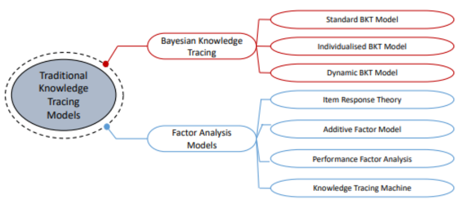

Fig.1. An overview of traditional knowledge tracing models

METHODOLOGY:

LPKT Model: Three modules make up each learning association's LPKT: the learning gain module, the failing to remember module, and the anticipating module. After the learner has finished an activity, the learning module shows the learning gains the student has made in comparison to earlier learning cooperation. The amount of information that will eventually be forgotten is predicted by the failing to remember module. The understudy's learning progress will next be assessed using the benefits of learning triumphs and the drawbacks of forgetting knowledge, updating his or her previous information state to acquire the most recent information state. Last but not least, it is recommended that the anticipating module forecast the student's presentation in the following task based on his or her most recent knowledge state.

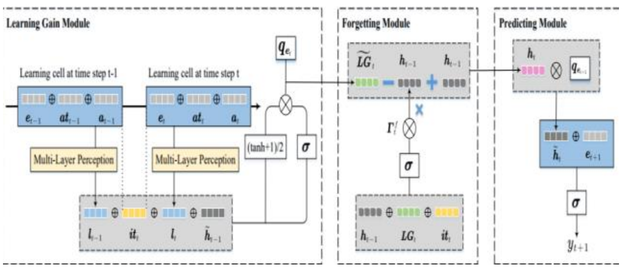

Fig.2. The architecture of the LPKT model

For ease of use, we only provide the handling at timestamp t and perform sporadic direct comparison calculations on the understudy's learning progression. The information is specifically the active learning cell et, at, its previous neighbour et1, att1, at1, and their time period. By monitoring the learning gain of the understudy and disregarding, LPKT determines the learning progress. Next, the student's knowledge state will be updated using the learning progress.

3.1 Learning Gain Module

The next step is to quantify the implied and dynamic learning progress in the growing experience because our main goal is to show the understudy progress for the KT task. This is done by formalising the growing experience as different mixtures of the fundamental learning cell and the time period.

$$NLG = \frac{post - pre}{1 - pre}, \qquad (1)$$

C Typically, a training or learning impact occurs when students respond to questions, which is the beneficial result of the learning gain. According to previous assessments, the learning increase is described as "distance travelled" [29], which refers to the variation in understudies' presentation at two distinct times. In light of this criteria, we should take into account the differences in how students presented themselves over two separate learning collaborations in order to clearly show the learning gain. In LPKT, we comprehend the exhibiting of learning gain by tying together students' prior learning to implant lt1 and current learning to install lt as the primary information component. However, despite the fact that we can detect differences in understudies' presentations using two continuous learning embeddings, it is unable to detect the diversity of learning acquired during the teaching process. For instance, not every student has the same learning experiences even though they present similarly in terms of the covered learning categories (i.e., have similar constant learning embeddings). The understudy's prior information state and the study duration are the next two factors we take into account when analyzing the learning gains. To completely combine the exercise embeddings, answer time embeddings, and answer embeddings, we concatenate et, at, and at and use a Multi-Layer Perceptron (MLP) as follows:

$$\boldsymbol{l}_t = \boldsymbol{W}_1^T[\boldsymbol{e}_t \oplus \boldsymbol{at}_t \oplus \breve{\boldsymbol{a}}_t] + \boldsymbol{b}_1, \qquad (2)$$

One point of view claims that the two-cell time frame, which reflects the qualities of learning gains, is a crucial component of the educational process. Understudies often learn more quickly and continuously because they are exposed to more material at shorter intervals. But the preceding informational state can also affect how well pupils learn; for instance, students with more pronounced dominance have more room for growth. In order to demonstrate the evolution of learning acquires, we consequently combine the above two parts into LPKT. Connect it specifically to the essential information component of the timeline in the course of events between the two consistent learning embeddings. To focus on the information state of the related information concepts of the current action for the previous information

ISSN 2347–3657
Volume 11,Issue 4, 2023

state, we The connected information state is obtained by first increasing ht1 and the information concept vector qet of the current action:

$$\widetilde{h}_{t-1} = q_{e_t} \cdot h_{t-1}, \qquad (3)$$

where the term "" stands for the inner product of two vectors. The learning gains will then be modelled using the following formula:

$$lg_t = tanh(W_2^T[l_{t-1} \oplus it_t \oplus l_t \oplus \widetilde{h}_{t-1}] + b_2), \qquad (4)$$

where W2 R (4dk) dk is the weight matrix, b2 R dk is the bias term, and tanh is the non-linear activation function. Because not all learning gains could be fully turned into the development of the students' knowledge, we also designed a learning trace (lt) to manage the students' absorptive limit of information. then it will be ready.

$$\Gamma_t^l = \sigma(W_3^T[l_{t-1} \oplus it_t \oplus l_t \oplus \widetilde{h}_{t-1}] + b_3), \qquad (5)$$

where R is the nonlinear sigmoid activation function, D is the bias term, W3 is the weight matrix, and b3 is the bias term.

The actual learning gain LGt in the twelfth learning contact is then calculated by multiplying l t by lgt. To expand the learning gain to other knowledge ideas, we increase LGt by qet to obtain LGgt, which stands for overall learning gains.

$$\begin{aligned} LG_t &= \Gamma_t^l \cdot ((lg_t + 1)/2), \\ \widetilde{LG}_t &= q_{e_t} \cdot LG_t, \end{aligned} \qquad (6)$$

Since the output range of the tanh function is (1, 1), we project the range of lgt from (1, 1) to (0, 1) using a linear transformation ((lgt + 1)/2). The learning gains LGt and LGgt will therefore always be positive, confirming our belief that students can continually learn throughout each learning engagement.

3.2 module of forgetting

The inverse forgetting peculiarity influences the amount of information that will be forgotten over time after being registered with LGgt, which improves pupils' information express. The amount of scholarly knowledge that is accumulated over time degrades significantly, according to the ignoring bend theory [34]. A basic manual-planned fantastic rot capacity is typically insufficient for identifying complex correlations between information state and temporal frame. In order to become familiar with the degree of loss data in information networks in light of three variables, we design an ignoring door ft in LPKT

ISSN 2347–3657
Volume 11,Issue 4, 2023

that makes use of an MLP.: (1) Understudies' current learning acquires LGt, (3) time frame, and (4) understudies' prior knowledge state ht1. This careless door foot in LPKT illustrates the subtle careless effects. Due to the significant non-linearity, a greater MLP is more equipped to identify the perplexing understudy's forgetting to recall behaviour in learning. The exact workout recipe is as follows:

$$\boldsymbol{\Gamma}_t^f = \sigma(\boldsymbol{W}_4^T[\boldsymbol{h}_{t-1} \oplus \boldsymbol{LG}_t \oplus \boldsymbol{it}_t]) + \boldsymbol{b}_4), \qquad (7)$$

The weight matrix is W4 R (3dk)dk, the non-linear sigmoid activation function is, and the bias term is b4 R dk.

The knowledge state of the understudy is then updated employing both the advantages of the learning gain and the drawbacks of not assessing the understudy's understanding advancement. Once students have attained the tth level of learning collaboration, the information state ht will be changed in the manner outlined below. To start with, we specifically duplicate ft to ht1 to counteract the impact of disregarding.

$$\begin{aligned} \boldsymbol{p}_t &= \widetilde{\boldsymbol{LG}}_t - \boldsymbol{\Gamma}_t^f \boldsymbol{h}_{t-1}, \\ \boldsymbol{h}_t &= \boldsymbol{p}_t + \boldsymbol{h}_{t-1}. \end{aligned} \qquad (8)$$

We also made an effort to update the information state via the brain network in the following ways because the brain network has demonstrated extraordinary ability to illustrate the non-linearity connection [41].

$$\boldsymbol{h}_t = \sigma(\boldsymbol{W}_N^T[\boldsymbol{h}_{t-1} \oplus \widetilde{\boldsymbol{LG}}_t \oplus \boldsymbol{\Gamma}_t^f]) + \boldsymbol{b}_N), \qquad (9)$$

The weight matrix is WN R (3dk)dk, while the bias term is bN R dk. As a result, to create the most up-to-date information state, the brain organization will mix the benefits of learning, the drawbacks of neglect, and the prior information state of the understudy. The effects of the brain mix in the tests, however, are marginally worse than those of the Eq expansion method.

3.3 Predicting Module: We have the latest information from the understudies right after the thirtieth learning communication by exhibiting their advancement in the educational process. In this section, we'll explain how to use ht to predict how well students will perform on the task that comes after this one (et+1).After reading another exercise et+1 in a real learning environment, the student will try to resolve it by using his or her insight to the contrasting information ideas. In this way, we infer the understudy's exhibition on et+1 using the associated information state het. A fully associated network with average activity and sigmoid initiation is used to project the initial link between het and the activity implanting et+1 to the result forecast.

$$y_{t+1} = \sigma(\frac{\sum(\boldsymbol{W}_5^T[\boldsymbol{e}_{t+1} \oplus \tilde{\boldsymbol{h}}_t] + \boldsymbol{b}_5)}{d_k}), \qquad (10)$$

International Journal of
Information Technology & Computer Engineering
ISSN 2347–3657
Volume 11,Issue 4, 2023

The weight matrix is W5 R (2dk) dk, while the bias term is b5 R dk. The output yt+1, which ranges in value from 0 to 1, displays the expected performance of the learner on the subsequent exercise, et+1. Additionally, if yt+1 is greater than the threshold, we may set a threshold that will let us know if a pupil has correctly answered et+1. If not, the answer is flawed.

3.4 Objective Function: The aim capacity to realize all boundaries in LPKT is also the cross entropy log difference between the predicted y and actual response an.

$$\mathbb{L}(\theta) = -\sum_{t=1}^{T}(a_t \log y_t + (1-a_t)\log(1-y_t)) + \lambda_\theta ||\theta||^2, \quad (11)$$

where stands for the regularization hyperparameter and signifies all LPKT parameters. The Adam optimizer [42] was applied to small batches to minimize the objective function.

THE LPKT-S MODEL: In order to determine how effectively pupils are learning, we plan to evaluate their learning progress, where their diverse learning rates have a big impact. In LPKT, understudies' explicit advancement rates are not recognized; instead, overall advancement rates are determined by comparing understudies' insight express, response time, and timespan. As a result, the rates of advancement across students with comparable informational states, response times, and time periods fundamentally vary. In this way, it is crucial to clearly define the rates at which students are learning. By introducing the understudy implanting with a single advancement rate for each understudy, we extend LPKT to LPKT-S in this section.

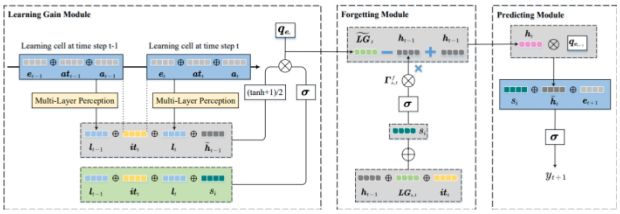

Fig.3. The architecture of the LPKT-S model

With three modules—learning gain, forgetting, and predicting—LPKT-S is comparable to LPKT. The distinction is that LPKT-S expressly assesses the impact of the student-specific development rate on the aforementioned three modules.

LPKTS modifies each of the three modules instead of LPKT. The information absorptive limit of students is specifically affected by the advancement rate in the learning gain module; in other words, students with faster progress rates are better at converting the underlying learning gain into their insight development. As a result, Eq. (4) also determines the understudy's underlying learning gain in LPKT-S, and while the understudy implanting si leaves the learning entryway l open, two persistent learning embeddings and their duration are shown below.

$$\boldsymbol{\Gamma}^{l}_{s,t} = \sigma(\boldsymbol{W}_6^T[\boldsymbol{l}_{t-1} \oplus \boldsymbol{it}_t \oplus \boldsymbol{l}_t \oplus \boldsymbol{s}_i] + \boldsymbol{b}_6), \qquad (12)$$

the weight structure is W6 R (3dk+ds)dk, and the inclination term is b6 R dk. Similar to LPKT, ls,t will then be copied to lgt at that time in order to obtain true learning LGs,t in the twelfth learning collaboration. Additionally, in order to obtain the general learning acquired by LGgs,t, we duplicate LGs,t by qet.

$$\begin{aligned} \boldsymbol{LG}_{s,t} &= \boldsymbol{\Gamma}^{l}_{s,t} \cdot ((\boldsymbol{lg}_t + 1)/2), \\ \widetilde{\boldsymbol{LG}}_{s,t} &= \boldsymbol{q}_{e_t} \cdot \boldsymbol{LG}_{s,t}. \end{aligned} \qquad (13)$$

In the forgetting module, we then add the student embedding to track how much information would be lost by various students because forgetting behaviours also vary among students.

As a result, in the failing to remember module, we further inform the student implanting with screen how much information will be failed to remember by various students in light of the first three components in LPKT. This is because the failing to remember ways of behaving are also variations among students. In LPKT-S, the neglecting entryway fs,t is handled as follows:

$$\boldsymbol{\Gamma}^{f}_{s,t} = \sigma(\boldsymbol{W}_7^T[\boldsymbol{h}_{t-1} \oplus \boldsymbol{LG}_{s,t} \oplus \boldsymbol{it}_t \oplus \boldsymbol{s}_i]) + \boldsymbol{b}_7), \qquad (14)$$

The weight matrix is W7 R (3dk+ds)dk, while the bias term is b7 R dk. We can also leverage the explicit benefits of learning for the understudy and the negative effects of failing to assess learning progress and update the understudy's information state. The computation technique is described as follows:

$$\begin{aligned} \boldsymbol{p}_{s,t} &= \widetilde{\boldsymbol{LG}}_{s,t} - \boldsymbol{\Gamma}^{f}_{s,t}\boldsymbol{h}_{t-1}, \\ \boldsymbol{h}_t &= \boldsymbol{p}_{s,t} + \boldsymbol{h}_{t-1}. \end{aligned} \qquad (15)$$

Finally, in the module on forecasting, we also take into account how understudy implantation affects understudies' ways of keeping track of their work. In the end, when using their knowledge to solve problems, understudies may possess a variety of traits. As a result, we add the student embedding si as follows to LPKT-S.

$$y_{t+1} = \sigma\left(\frac{\sum(\boldsymbol{W}_8^T[\boldsymbol{e}_{t+1} \oplus \widetilde{\boldsymbol{h}}_t \oplus \boldsymbol{s}_i] + \boldsymbol{b}_8)}{d_k}\right), \qquad (16)$$

The weight matrix is W8 R (2dk+ds)dk, while the bias term is b8 R dk. Then, by restricting a similar target capability, we can get ready for LPKT-S. We are able to comprehend the complex LPKT-S model, which explicitly acknowledges the varying rates of development of understudies, by bringing the understudy implanting into the three LPKT modules.

RESULTS

The Results after testing LPKT Model

ISSN 2347–3657
Volume 11,Issue 4, 2023

```
PS C:\Users\ddivy\Desktop\New folder\LPKT-S-main> python test.py 1691298021
(400, 7)
runs/1691298021/bestcheckpoints/best_checkpoints
epochs 0: rmse 0.409516  auc 0.771625  r2 0.206174  acc 0.751948
```

The Results after testing LPKT-S Model

```
PS C:\Users\ddivy\Desktop\New folder\LPKT-S-main> python test.py 1691332818
(400, 7)
runs/1691332818/bestcheckpoints/best_checkpoints
epochs 0: rmse 0.405566  auc 0.780336  r2 0.221414  acc 0.757181
```

Summary based on the results of two models:

The key difference we can see in this two models is the auc and r2 evaluation metrics are higher in lpkt model compare with lpkt-s model.

CONCLUSION:

First, a tuple made up of exercise, response time, and answer was used to characterise the fundamental learning cell. The learning method was then formalised by combining interval times and fundamental learning cells. We then created a model that depicted the positive impact of learning gain and the detrimental impact of forgetting in the learning process in order to monitor students' progress and update their knowledge states. Considering that pupils often progress at different speeds. In our work, we employ the KT task, which yields more informative outcomes for boosting learning and teaching, as a viable future research topic.

FUTURE SCOPE:

We'll keep looking towards better ways to assess students' learning progress in the future. For example, we may incorporate what teachers and students have said about their own teaching strategies. We can also look into the idea that objective function constraints could speed up model learning. Additionally, we'll consider gauging student growth rates at the level of knowledge concepts in order to more precisely quantify student progress rates. Finally, we will look into how to precisely describe the relationship between workouts and knowledge concepts by automatically learning specific weights in the Q-matrix.

REFERENCE:

1] Ebba Ossian Nilsson. Sustainability: Special issues "the futures of education in the global context: Sustainable distance education". *Sustainability*, 07 2020.

[2] Tuan Nguyen. The effectiveness of online learning: Beyond no significant difference and future horizons. *MERLOT Journal of Online Learning and Teaching*, 11(2):309–319, 2015.

[3] Cristobal Romero and Sebasti ´ an Ventura. Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(6):601–618, 2010.

[4] Albert T Corbett and John R Anderson. Knowledge tracing: Modelling the acquisition of procedural knowledge. *UMUAI*, 4 (4):253–278, 1994.

[5] Radek Pelanek. Bayesian knowledge tracing, logistic models, ´ and beyond: an overview of learner modelling techniques. *User Modeling and User-Adapted Interaction*, 27(3-5):313–350, 2017.

[6] Le Wu, Yong Ge, Qi Liu, Enhong Chen, Richang Hong, Junping Du, and Meng Wang. Modelling the evolution of users' preferences and social links in social networking services. *IEEE Transactions on Knowledge and Data Engineering*, 29(6):1240–1253, 2017.

[7] Dongmin Shin, Yugeun Shim, Hangyeol Yu, Seewoo Lee, Byung soo Kim, and Youngduck Choi. Saint+: Integrating temporal features for ednet correctness prediction. *arXiv preprint arXiv:2010.12042*, 2020.

[8] Zhenya Huang, Xin Lin, Hao Wang, Qi Liu, Enhong Chen, Jianhui Ma, Yu Su, and Wei Tong. Disenqnet: Disentangled representation learning for educational questions. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 696–704, 2021.

[9] Ting Long, Jiarui Qin, Jian Shen, Weinan Zhang, Wei Xia, Ruiming Tang, Xiuqiang He, and Yong Yu. Improving knowledge tracing with collaborative information. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 599– 607, 2022.

[10] Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, and Yonghe Zheng. A survey of knowledge tracing. *arXiv preprint arXiv:2105.15106*, 2021.

[11] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[12] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. In *NeurIPS*, pages 505–513, 2015.

[13] Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, and Guoping Hu. Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, 33(1):100–115, 2019.

[14] Shuanghong Shen, Qi Liu, Enhong Chen, Zhenya Huang, Wei Huang, Yu Yin, Yu Su, and Shijin Wang. Learning process consistent knowledge tracing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1452–1460, 2021.

[15] Julia KaFer, Susanne Kuger, Eckhard Klieme, and Mareike Kunter. . The significance of dealing with mistakes for student achievement and motivation: results of doubly latent multilevel analyses. *European Journal of Psychology of Education*, 2019.

ISSN 2347–3657
Volume 11,Issue 4, 2023

[16] Gabriele Steuer and Markus Dresel. A constructive error climate as an element of effective learning environments. *Psychological Test and Assessment Modeling*, 57(2):262 – 275, 2015.

[17] Safer, Nancy, Fleischman, and Steve. How student progress monitoring improves instruction. *Educational Leadership*, 2005. [18] K. Mclane. What is student progress monitoring and how will it help me?. *National Center on Student Progress Monitoring*, page 4, 2006.

[19] Ye Mao. Deep learning vs. bayesian knowledge tracing: Student models for interventions. *Journal of educational data mining*, 10(2), 2018.

[20] Shaul Markovitch and Paul D Scott. The role of forgetting in learning. In *Machine Learning Proceedings 1988*, pages 459–465. Elsevier, 1988.

[21] Philip I Pavlik Jr, Hao Cen, and Kenneth R Koedinger. Performance factors analysis–a new alternative to knowledge tracing. *Online Submission*, 2009.

[22] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. Dynamic key-value memory networks for knowledge tracing. In *WWW*, pages 765–774, 2017.