U N I V E R S I T Y O F S P L I T
FACULTY OF ELECTRICAL ENGINEERING, MECHANICAL ENGINEERING
AND NAVAL ARCHITECTURE

Ines Šarić-Grgić

ADAPTIVE FORMATIVE ASSESSMENT BASED ON ENHANCED BAYESIAN KNOWLEDGE TRACING MODEL

DOCTORAL THESIS

Split, 2025.

U N I V E R S I T Y O F S P L I T
FACULTY OF ELECTRICAL ENGINEERING, MECHANICAL ENGINEERING
AND NAVAL ARCHITECTURE

Ines Šarić-Grgić

Adaptive Formative Assessment Based on Enhanced Bayesian Knowledge Tracing Model

DOCTORAL THESIS

Split, 2025.

The research reported in this thesis was carried out at (Department name), University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture.

Supervisor: Prof. dr. sc. Ani Grubišić, Faculty of Science, University of Split, Croatia
Disertation number: xxx-xxx

BIBLIOGRAPHIC INFORMATION

Keywords: Bayesian knowledge tracing, student modelling, educational data mining, intelligent tutoring systems
Scientific area: Technical Sciences
Scientific field: Computer Science
Scientific branch: Artificial Intelligence
Institution of PhD completion: University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
Supervisor of the thesis: Prof. dr. sc. Ani Grubišić
Number of pages: 182
Number of figures: 32
Number of tables: 172
Number of references: 131

Committee for assessment of doctoral dissertation:

1. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
2. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
3. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
4. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
5. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town

Committee for defence of doctoral dissertation:

1. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
2. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
3. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
4. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town
5. Xxxx. dr. sc. Xxxx Yyyyy, Institution name and City/Town

Dissertation defended on: xx. month. 20xx.

Adaptive Formative Assessment Based on Enhanced Bayesian Knowledge Tracing Model

Abstract:

This doctoral thesis utilised enhanced Bayesian Knowledge Tracing (BKT) models to assess student knowledge adaptively and to extend the standard functionalities of widely used Learning Management Systems (LMS). It was hypothesised that incorporating time spent on task and the number of code evaluations in the introductory programming domain would enhance student performance's predictive accuracy and knowledge mastery estimation within the vanilla BKT model. Empirical research employing in situ quasi-experimental design was conducted during an Introduction to Programming in Python course, involving a substantial sample of 174 undergraduate students. The weekly course topics were structured into granular Domain Knowledge Components (DKC) and Controlled Environment formative assessments (CE). The study examined 18 BKT models incorporating various combinations of Prior knowledge, Guess, Slip, Learn and Forgets parameter probabilities, implemented using the Python library for cognitive modelling, pyBKT. In general, the enhanced BKT models outperformed the baseline vanilla model in predicting student performance across multiple DKCs. Also, the enhanced BKT models outperformed the vanilla model in estimating students' knowledge mastery. Regarding the model convergence, the enhanced BKT models provided more effective and reliable paths to knowledge mastery than the vanilla model. The previous results supported the proposal of a framework for ranking BKT models based on their capacity to predict student performance, estimate knowledge mastery, and model efficient learning paths. This framework identified the most effective BKT models, offering a systematic approach to selecting models that outperformed the vanilla BKT model.

Keywords:
Bayesian knowledge tracing, student modelling, educational data mining, intelligent tutoring systems

Prilagodljivo formativno vrednovanje temeljeno na poboljšanom modelu Bayesovog praćenja znanja

Sažetak:

Ne temelju prethodnih poboljšanja osnovnog Bayesovog modela za praćenje znanja učenika (eng. Bayesian Knowledge Tracing, BKT), u radu se istražuju BKT modeli za prilagodljivu procjenu znanja, proširujući pritom standardne funkcionalnosti sustava za e-učenje. Postavljena je hipoteza da uvođenje značajki vremena provedenog na pitanju i broja evaluacija koda u području početnog programiranja poboljšava prediktivnu točnost učenikovih odgovora te procjenu razine usvojenog znanja. Empirijsko istraživanje provedeno je korištenjem in situ kvazi-eksperimentalnog dizajna tijekom kolegija Uvod u programiranje u kojem je sudjelovalo 174 studenta prijediplomskog studija. Tjedni nastavni sadržaji strukturirani su u granularne komponente znanja (eng. Domain Knowledge Components, DKC) uz pripadajuća formativna vrednovanja (eng. Controlled Environment assessments, CE). Korištenjem pyBKT Python biblioteke za kognitivno modeliranje, istraženo je 18 BKT modela koji uključuju različite kombinacije parametara predznanja (eng. Prior), pogađanja (eng. Guess), slučajne pogreške (eng. Slip), učenja (eng. Learn) i zaboravljanja (eng. Forgets). Predloženi BKT modeli nadmašili su osnovni model u predviđanju uspjeha učenika u više DKC-ova. Također, nadmašili su osnovni model u procjeni razine usvojenog znanja. U kontekstu konvergencije modela, predloženi BKT modeli pružili su učinkovitije i pouzdanije modeliranje individualnih pristupa učenju. Rezultati istraživanja rezultirali su prijedlogom okvira za rangiranje BKT modela s obzirom na njihovu sposobnost predviđanja učenikovih odgovora, procjene razine usvojenog znanja i modeliranje individualnih pristupa učenju. Predloženi okvir identificirao je najučinkovitije BKT modele, nudeći sustavan pristup odabiru modela koji nadmašuju rezultate osnovnog BKT modela.

Ključne riječi:
Bayesov model praćenja znanja, modeliranje učenikovog znanja, rudarenje podataka u obrazovanju, inteligentni tutorski sustavi

Acknowledgments

Contents

List of Tables

List of Figures

1 INTRODUCTION

Educational Data Mining (EDM) is a progressive area of scientific research focused on developing methodologies for analysing unique data collected from educational environments to gain deeper insights into students and their learning contexts [7]. While closely related to Learning Analytics (LA), EDM emphasises automated methods over human interpretation of data and visualisation [8]. Researchers in EDM employ various techniques, including data mining to identify patterns, machine learning to glean insights from training data and predict future outcomes, and statistical analysis to quantify data from samples and estimate population behaviour. A key moment in the evolution of the EDM field was the publication by Corbett and Anderson on the Bayesian Knowledge Tracing (BKT) model [8, 9], which marked the first significant milestone.

Although educational data is not exclusively sourced from digital platforms, e-learning systems represent a broad testing ground. Commonly referred to as Learning Management Systems (LMS) [10], these systems are designed to manage, document, track, report, automate, and deliver courses, materials and learning experiences. However, the scope of e-learning systems extends beyond LMS, encompassing any educational system that employs formalised teaching supplemented by electronic resources such as computers and the Internet. Additionally, another significant application of e-learning systems is the delivery of Massive Open Online Courses (MOOC). Despite their widespread use, neither educational environment was originally designed to provide adaptive and intelligent functionalities.

Since the 1960s, researchers in the interdisciplinary field of cognitive science, artificial intelligence, and educational technology have computerised teaching and learning by developing various types of adaptive educational environments, e.g. Computer Assisted Instruction (CAI) systems [11], Intelligent Tutoring Systems (ITS) [12], Intelligent Learning Environments (ILE) [13], Adaptive Instructional Systems (AIS) [14], etc. One of the primary strengths of these environments is their ability to track and analyse student knowledge and behaviour, thereby enabling them to identify each student's individual needs. The effectiveness of other components within these platforms heavily depends on the student model's capacity to represent student knowledge accurately. The extensive research conducted in the field of ITSs over the past few decades represents a valuable resource, showcasing various approaches to student modelling. While the precise taxonomy of these approaches remains a subject of ongoing debate, Machine Learning (ML) techniques have gradually emerged as the cornerstone for categorising student modelling approaches [1].

According to previous review studies on student modelling [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 1, 26, 27], one of the earliest and most extensively investigated ML approaches is the BKT model [9]. This model, founded on Hidden Markov Models, exemplifies the current state of the art in the field. A significant advantage of the BKT approach is its flexibility in accommodating the limited data sets often encountered in typical class sizes.

The vanilla BKT model refers to the original and basic version of the model that emerged from the development efforts of the ACT Programming Tutor and the cognitive theory referred to as ACT-R (Adaptive Control of Thought–Rational) [28]. According to the ACT-R theory, mastering a complex skill requires the mastery of its individual components. The vanilla BKT model also adopts Bayesian computation principles from Atkinson's work [29]. It conceptualises the student's knowledge mastery as a latent variable using the Hidden Markov Model (HMM) as a specific type of Bayesian network. The HMM has nodes representing knowledge states (learned or unlearned) and performance states (correct or incorrect). Transitions between these nodes are determined by probability parameters—such as prior knowledge, guessing, slipping, and learning—specified by experts for each skill. Specifically, the transition from a learned to an unlearned state is absent in the vanilla BKT model, aligning with the "no forgetting" paradigm. Additionally, the model considers independent knowledge components, comprising sets of questions that are of equal difficulty. This approach also accommodates a student's initial attempts, allowing for multiple tries at answering a question.

Since its introduction, researchers have explored numerous enhancements to the vanilla BKT model. These enhancements have been evaluated based on architectural features, educational context, and extensions not accounted for in the vanilla BKT model. The most extensively investigated enhancements include student characteristics and tutor interventions [1]. Typically, researchers have explored the application of BKT within educational settings such as ITS, which incorporate both instructional and assessment components. However, the use of BKT in MOOC or simulated environments has been comparatively rare. Enhanced BKT models have primarily been evaluated based on their predictive capabilities regarding the accuracy of students' answers. In contrast, only a limited number of studies have focused on their effectiveness in estimating knowledge mastery [30, 31, 32, 33, 34, 35]. Enhanced BKT models have been extensively studied in mathematics education, while comparatively fewer approaches have been applied in programming and language learning contexts.

In the programming domain, the educational environment explored in the vanilla BKT research was centred around the ACT Programming Tutor, which facilitated the practice of short programs in languages such as Lisp, Prolog, and Pascal. González-Brenes et al. [36] introduced the Feature-aware Student Knowledge Tracing method and proposed an enhanced BKT architecture, which incorporates performance node features such as subskills, item difficulty, and the number of answer opportunities. Similarly, Khajah et al. [37] investigated the

QuizJET Java Programming Tutor and proposed a model enhanced by the Item Response Theory, which enabled the modelling of various student abilities and problem difficulties. Huang et al. [38, 39] used the SQL-KNOT and JavaGuide platforms to investigate skill combinations that might involve additional specific knowledge. Wang et al. [40] employed a platform for teaching the C programming language and introduced an enhanced model that integrates the 2PL Item Response Theory, which estimates students' prior knowledge and combines it with the discrimination and difficulty levels of each assessed skill.

Since the 1960s, the Association for Computing Machinery (ACM) and the wider community have collaborated to formulate standards and guidelines for Computer Science (CS) curricula [41, 42]. Their recent work on the curricular guidelines for the upcoming decade resulted in the initial draft of the Future of CS educational materials [42]. In light of prevailing trends, educators, administrators, authors, and practitioners have highlighted the significance of educational environments such as ITS and LMS and the system's capacity for auto-graded assessment. They have acknowledged the pioneering role of CS education in creating interactive learning content, which includes algorithm animations and program visualisations [43], programming problems with auto-grading [44] and ITS [45]. Furthermore, they have emphasized that utilising a pool of auto-graded questions enables students to engage in regular self-assessment at various stages of the learning process—when they encounter new information (e.g. through reading), apply it (e.g. by completing homework or in-class activities), and review it (e.g. during summative assessments). The objective is to minimise the gap between students' willingness to practice and their actual ability to do so, thereby enhancing the efficiency of each learning hour spent on a task [42]. Overall, auto-grading increases the adaptability of educational platforms by reducing the number of questions assigned to students who answer correctly.

Pelánek [46] presented an overview of the terminology typically used in the literature to define three levels of adaptivity prevalent in educational environments. The first level pertains to adaptation within a single item (also referred to as a task, problem, or question), which typically entails tailoring various types of learning support—such as hints, scaffoldings, feedback, and explanations—to meet students' individual needs. In the literature, this level is also known as the inner loop [47], micro-adaptation [48], or step loop [49]. The second level involves adaptation within broader instructional steps, such as selecting or recommending exercises and topics for study. This level is also known as the outer loop [47], macro-adaptation [48], or task loop [49]. The third level involves adapting the educational system itself, such as adding or removing items or modifying algorithmic settings, with the adaptation process being either automated or supervised by a human. This level is commonly known as the design loop [49], closing-the-loop [50] or human-in-the-loop [23]. This thesis investigates assessment as a dynamic and personalised approach to evaluating a student's knowledge, aligning with the first adaptivity level. In the educational setting, assessments serve multiple purposes, including feedback, fostering learning through practical application and acting as external motivators [42]. Research has indicated that frequent assessments can improve student performance and alleviate test anxiety [51, 42, 52], as well as allow students to retake exams if they have not fully mastered the material [42, 53, 54, 55]. Formative assessment primarily targets student learning, focusing on three key components: productive time spent on tasks, feedback, and spaced repetition [42].

The primary motivation behind this thesis is to enhance the BKT model for the adaptive assessment of student knowledge. Although this functionality is typically a key feature of student modelling in adaptive educational systems such as ITSs, the approach here extends beyond the standard functionalities of LMSs. The approach focuses on formative assessments, utilizing auto-graded tasks to allow teachers to determine the minimum time and number of questions required to achieve knowledge mastery. The flexibility of BKT, especially in managing the limited data sets typical of standard class sizes, represents a significant advantage, particularly in larger-scale educational environments where time efficiency is critical. The empirical research employed an in situ quasi-experimental design within an introductory programming course involving a sample of 174 students. To ensure the acquisition of the fine-grained domain knowledge required for applying the BKT model, the approach was designed around formative assessments aligned with weekly course topics. In addition to the time spent on each programming question, the approach considers how often a student checks the question's correctness using pre-defined test cases. By classifying student answers based on these features, a more targeted assessment framework that defines specific BKT parameter probabilities tailored to the context of this research is proposed. Enhanced BKT modelling was prioritized for adaptive assessment of student knowledge, with minimized influence from other factors by using questions of equal difficulty and randomising question sequences.

1.1 Research hypotheses

In response to the identified motivational challenges, this thesis aims to explore the enhanced BKT modelling for adaptive assessment. Accordingly, it addresses two primary hypotheses:

Hypothesis 1

It is feasible to trace student knowledge using Bayesian modelling and leveraging the time spent on task and the number of code evaluations.

The proposed hypothesis is investigated through a series of research tasks, including data collection, data pre-processing, BKT parameter fitting, and evaluation of the BKT models.

Hypothesis 2

Extending the modeling of student knowledge with features such as the time spent on task and the number of code evaluations enhances the vanilla BKT model.

This thesis presents the framework for the BKT model ranking based on its effectiveness in predicting student performance and accurately estimating knowledge mastery.

1.2 Dissertation outline

This doctoral thesis comprises six chapters. Chapter 1 introduces the research motivation and outlines the study's hypotheses. Chapter 2 presents an overview of the theoretical foundations of the vanilla BKT model and systematically reviews its enhancements.

Chapter 3 describes the research methodology in detail, encompassing data collection, data preprocessing, BKT parameter fitting and BKT model evaluation. It also outlines the methodological instruments employed to address the research hypotheses. Chapter 4 presents and discusses the results, including an analysis of formative assessment data, classification of student answers based on time on task and the number of code evaluations, BKT model parameters, student performance predictions, knowledge mastery estimations and student learning paths. This chapter also ranks the examined BKT models and provides a general discussion of the findings. Chapter 5 demonstrates the BKT-based adaptive formative assessment by detailing the BKT Quiz Report prototype module, BKT API, BKT Quiz prototype module and the experimental guidelines.

Finally, Chapter 6 concludes the thesis by summarizing the findings, emphasizing the scientific contributions and discussing potential directions for future research.

2 RELATED WORK

This chapter draws extensively from the systematic review of enhanced BKT modelling approaches [1]. Section 2.1 reviews previous studies on student modelling, while Section 2.2 presents a brief theoretical overview of the BKT model. Finally, Section 2.3 details a systematic review of enhanced BKT modelling approaches, employing the PRISMA methodology to rigorously select and analyze relevant studies.

2.1 Student modelling

The aim is to present (i) the application of BKT in the previous review studies on student modelling, (ii) an overview of ML techniques used for student modelling, and (iii) a comparison of the Bayesian Network-based BKT model with Logistic Regression-based and Neural Network-based models [1].

The summary of previous literature on student modelling draws from overviews [16, 18, 19, 20, 26], reviews [15, 21, 22, 25, 27, 23] and systematic literature reviews [17, 24]. Table 2.1 provides a comprehensive list of proposed taxonomies of student modelling approaches in descending chronological order, including references, research focus and the proposed taxonomy. The BKT-related student modelling approaches are indicated in italics.

Previous research on student modelling approaches primarily originated within the ITS and adaptive instruction research communities. Acknowledgement is given to the extensive body of work in cognitive diagnostic modelling and psychometrics stemming from other research fields, including Item Response Theory, the DINA model family and related frameworks [56]. However, these approaches are not included in the present review of ML methods for student modelling.

Each study in Table 2.1 provides an overview of student modelling approaches related to specific educational platforms, adaptive behaviour and techniques used. In the early reviews, ML techniques were already the basis for different student modelling approaches. Over time, new techniques complemented the previous student modelling taxonomies. However, the new taxonomies of student modelling approaches are still adopted, and there is no consensus on the correct taxonomy.

As the subfield of artificial intelligence, ML works on algorithms that enable machines to learn through experience and data [57]. ML techniques used for student modelling en-

Table 2.1. Proposed taxonomies of student modelling approaches [1].

Reference	Research focus	Proposed taxonomy of student modelling approaches
Liu et al. (2021)	The review from the technical point of view	*Probabilistic models;* Logistic models; Deep learning-based models
Ramirez Luelmo et al. (2021)	The systematic review of machine learning techniques (2015-2020)	*Bayesian Knowledge Tracing;* Deep Knowledge Tracing; Long-Short Term Neural Networks; Bayesian Networks; Support Vector Machines; Dynamic Key-Value Memory Network; Performance Vector Analysis
Pelanek (2017)	The review focused on the macro adaptive behaviour (curriculum sequencing) (2014-2020)	*Bayesian Knowledge Tracing;* Logistic models
Anouar Tadlaoui et al. (2016)	The review focused on Adaptive Educational Systems	Overlay; Stereotype; Machine Learning*;* Plan Recognition; Differential; Perturbation; *Bayesian Networks*
Sani et al. (2016)	The review focused on Intelligent Tutoring Systems (2010-2015)	*Bayesian Knowledge Tracing;* Fuzzy Logic; Overlay; Differential; Perturbation; Constraint-based; Machine Learning; Stereotype
Kurup et al. (2016)	The overview focused on Intelligent Tutoring Systems	Overlay; Bayesian Network; Correct First Attempt Rate; Performance Factor Analysis; Tabling; *Bayesian Knowledge Tracing*
Zafar & Ahmad (2013)	The review of the student modelling approaches under uncertain conditions	Student modelling using statistical reasoning (*Bayesian Networks*, Reasoning using Certainty Factors); Fuzzy Modelling
Pavlik et al. (2013)	The review focused on Intelligent Tutoring Systems	Overlay models (Rule Space models, *Model Tracing models,* Constraint-based models); Knowledge Space models; Dialogue models; Programmed Branching; State and Trait
Chrysafiadi & Virvou (2013)	The systematic review focused on Adaptive Educational Systems (2002-2012)	Overlay; Stereotypes; Perturbation; Machine Learning; Cognitive Theories; Constraint-based Model; Fuzzy Modelling; *Bayesian Networks*; Ontology-based Modelling
Harrison & Roberts (2012)	The overview of student modelling techniques for application in serious games	*Knowledge Tracing;* Performance Factor Analysis; Matrix Factorization
Desmarais & Baker (2012)	The overview of the most successful and widely used approaches focused on the macro adaptive behaviour (curriculum sequencing)	Tutors for problem-solving and solution analysis (Cognitive tutors and Constraint-based modelling); Content sequencing tutors (Models of skills - Bayesian Networks and graphical models, IRT and Latent Trait models, Latent cousins DINA, NIDA, DINO, NIDO, *Bayesian Knowledge Tracing*, Models without hidden nodes)
Vandewaetere et al. (2011)	The overview of the parameters that are included in the student model when developing adaptive learning environments	Stereotypes; Feature-based modelling; Combination of stereotypes and feature-based modelling; Other approaches (Constraint-based modelling and Modelling of misconceptions)
Brusilovsky & Milan (2007)	The overview focused on Adaptive Hypermedia and Adaptive Educational Systems	Overlay; *Uncertainty-based modelling*

Table 2.2. ML techniques identified in the research on student modelling [1].

Research study	Bayesian Networks	Logistic Regression	Neural Networks	Support Vector Machines	Fuzzy Logic	Matrix Factorization
Liu et al. (2021)	+	+	+	-	-	-
Ramirez Luelmo et al. (2021)	+	+	+	+	-	-
Pelanek (2017)	+	+	-	-	-	-
Anouar Tadlaoui et al. (2016)	+	-	-	-	-	-
Sani et al. (2016)	+	-	-	-	+	-
Kurup et al. (2016)	+	+	-	-	-	-
Zafar & Ahmad (2013)	+	-	-	-	+	-
Pavlik et al. (2013)	+	-	-	-	-	-
Chrysafiadi & Virvou (2013)	+	-	-	-	+	-
Harrison & Roberts (2012)	+	+	-	-	-	+
Desmarais & Baker (2012)	+	+	-	-	-	-
Vandewaetere et al. (2011)	-	-	-	-	-	-
Brusilovsky & Milan (2007)	+	-	-	-	+	-

hance the adaptiveness and intelligence of educational platforms. Those identified in the already mentioned research studies include Bayesian Networks, Logistic Regression, Neural Networks, Support Vector Machines, Fuzzy Logic, and Matrix Factorization (Table 2.2).

Moreover, Liu et al. [21] classified the student modelling approaches as Probabilistic, Logistic, and Deep Learning-based models. Probabilistic models, such as BKT, are based on

Bayesian Networks and assume that the learning process follows a Markov process, which uses the observed states to estimate the student's hidden knowledge states. Logistic models, such as Learning Factor Analysis [58] and Performance Factor Analysis [59], predict the probability of student performance by learning function, typically a logistic function. The last group of knowledge tracing approaches are Deep Learning models based on Neural Networks [60]. The shortest high-level classification of student modelling approaches is supported, though it presents some inconsistency, as logistic models may also be considered probabilistic.

Ramirez Luelmo et al. [24] investigated ML techniques employed in student modelling from 2015 to 2020. Their research results indicate the most common ML techniques as BKT (18 applications), Deep Knowledge Tracing (13 app.), Long-Short Term Neural Networks (12 app.), Bayesian Networks (11 app.), Support Vector Machines (7 app.), Dynamic Key-Value Memory Networks (7 app.), and Performance Factor Analysis (6 app.).

Overall, Bayesian Networks are the continuously investigated ML technique used for student modelling. The vanilla BKT based on the Hidden Markov Model is the most representative and unique student modelling approach researchers consider a baseline.

As for the preference between the probabilistic Bayesian Network-based and Logistic Regression-based models, researchers often prefer one model but provide no rationale behind their choices [23]. On the other side, the apparent accuracy improvement of Deep Learning-based models over BKT was due to the high dimensional hidden space and ability to observe interleaved skills in a single model [61]. The comparison between Neural Network-based research and the vanilla BKT model revealed that simply enabling the forgetting parameter of the vanilla model led to a performance close to Deep Knowledge Tracing on several datasets [62, 63]. Based on Bayesian statistics, the BKT model assumes nodes with binary states and is more interpretable than the Neural Network-based model.

2.2 Bayesian Knowledge Tracing (BKT) model

A Bayesian Network is a probabilistic graphical model for representing knowledge about an uncertain domain, where each node represents a random variable and directed edges between nodes indicate probabilistic dependencies between these variables. A Hidden Markov Model (HMM) is a special Bayesian Network for tracing not directly observable (hidden) nodes using the observable node states. In BKT, hidden nodes represent student knowledge, and observable nodes represent student performance. Both nodes are assumed to be binary, including the unlearned and learned knowledge states and the correct and incorrect performance states.

Figure 2.1 (based on Zhang and Yao [1, 2]) shows the hidden student $knowledge_t, t \in \{1,2,...T\}$ and observable student $performance_t nodes, t \in \{1,2,...T\}$ of the vanilla HMM. While $P(L_0)$ is the initial probability of knowledge before any opportunity to apply it (Prior),

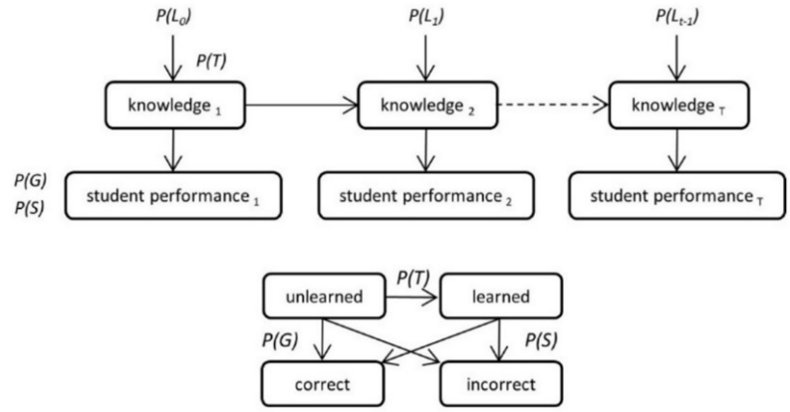

Figure 2.1. The vanilla BKT model and its instantiation process - based on Zhang and Yao [2, 1].

Priors	
learned	$P(L_0)$
unlearned	1- $P(L_0)$

Transitions		
	to learned	to unlearned
from learned	1	0
from unlearned	$P(T)$	1- $P(T)$

Observations		
	correct	incorrect
learned	1- $P(S)$	$P(S)$
unlearned	$P(G)$	1- $P(G)$

Figure 2.2. BKT parameters in a matrix form [3].

there are also transition and emission probabilities. The transition probabilities refer to the probability P(T) of knowledge transitioning from an unlearned state to a learned state (Learn) and to the probability P(F) of forgetting a previously known knowledge, which is assumed to equal zero in the vanilla model (Forgets). The model defines emission probabilities by guessing the probability of correctly answering unlearned knowledge P(G) (Guess) and the slip probability of making a mistake when answering a learned knowledge P(S) (Slip). Figure 2.2 shows the complete set of vanilla model parameters consisting of Prior $P(L_0)$, Learn P(T), Guess P(G), and Slip P(S) in a matrix form.

The main task of the vanilla model is to estimate the probability that a student has mastered the knowledge at time step t, denoted by a learning parameter $P(Lt), t \geq 0$. The model updates the probability $P(Lt)$ after each opportunity to apply knowledge given an observed correct or incorrect response as follows:

$$P(L_{t-1}|Correct_t) = \frac{P(L_{t-1})(1-P(S))}{P(L_{t-1})(1-P(S))+(1-P(L_{t-1}))P(G)}$$

$$P(L_{t-1}|Incorrect_t) = \frac{P(L_{t-1})P(S)}{P(L_{t-1})P(S)+(1-P(L_{t-1}))(1-P(G))}$$

If $evidence_t \in \{Correct_t, Incorrect_t\}$ represents the observable correctness of a student's answer after an opportunity t to apply knowledge, the updated probability for the following time step is defined as:

$$P(L_t) = P(L_{t-1}|evidence_t) + (1 - P(L_{t-1}|evidence_t))P(T)$$

Using the evidence from the current step, the model first calculates the probability that the student knew the answer before making an attempt. Then, taking this into account, it computes the likelihood that the student learned it after making the attempt. The previous equations are based on the original BKT publication [9], and a similar notation appears in the work by Zhang and Yao [2].

Regarding the BKT parameter estimation procedure, Corbett and Anderson [9] discussed individualisation per skill and individualisation per student of all four BKT parameters. The individualised BKT model resulted in a better correlation between actual and expected accuracy across student results than the non-individualized BKT model, whose accuracy of predicting student test scores (after working with a tutoring system) did not improve tangibly [3]. Finally, the vanilla model's parameter fitting procedure involves expert-based estimations of the four BKT parameters per skill.

2.3 Systematic review of enhanced BKT modelling

This work differs from other literature reviews on several accounts since it focuses on the probabilistic BKT models, systematically covers the research works published since the introduction of BKT in 1995 up to the most recent research in 2022, and reviews the BKT enhancements and evaluation approaches, including datasets from educational platforms and the performance measures found in the literature. It is built upon the framework of ITS and the vanilla BKT model, which is recognized as one of the first ML and most representative approaches. We aim to discuss two Research Questions (RQ):
RQ1: What has been proposed in the literature to enhance the vanilla BKT model since its emergence in 1995?
RQ2: Which evaluation approaches, including data collected from educational platforms and performance measures, were part of the research on the BKT enhancements?

2.3.1 Methodology

The methodology used in this work is in line with the PRISMA guidelines [64] consisting of (i) Rationale, objectives and research questions, (ii) Eligibility criteria, information sources, and a search strategy, (iii) Screening process and study selection and (iv) Data collection and features. Since we have elaborated on the rationale and objectives in the previous sections, we proceed with the criteria, sources, and search strategy for works that fall under the scope of this review.

The main eligibility criterion referred to scientific works on the vanilla BKT model enhancements published in the relevant scientific databases until 2022. The implementations

Table 2.3. Database search details [1].

Database	Search query (- 2022)	#
Web of Science	"knowledge tracing" (Abstract) AND (bayes* OR probab*) (Abstract)	89
Scopus	ABS ("knowledge tracing") AND ABS ((bayes* OR probab*))	200
ACM	[Abstract: "knowledge tracing"] AND [[Abstract: bayes*] OR [Abstract: probab*]]	25
IEEE Xplore	("Abstract":"knowledge tracing") AND ("Abstract":bayes* OR "Abstract":probab*)	20
Google Scholar	allintitle: "bayesian knowledge tracing"	75

of the BKT enhancements could proceed in two directions, including the Bayesian network architecture/educational context and new computational methods.

We searched the scientific databases indexing quality-proven journals and conference proceedings, including the Web of Science (Core Collection), Scopus, ACM (Full-Text Collection), IEEE Xplore, and Google Scholar (the final refinements made in March 2023). The search strategy included the expression knowledge tracing and versions of Bayes and probabilistic words contained in the publication abstracts. Due to the extensiveness of the Google Scholar database, we searched the publication titles using the expression Bayesian knowledge tracing. Table 2.3 shows the search details with the number of publications.

Figure 2.3 shows the PRISMA flow diagram of the publication identification and screening process, a widely accepted method for conducting systematic reviews. It visually represents the steps taken to identify and screen publications, ensuring transparency and reproducibility in the review process.

Out of 409 results from the five academic databases, we compiled 223 publications (177 duplicates and nine conference proceedings removed).

The screening of abstracts resulted in 84 excluded publications, which were off-topic and written in languages other than English or as a programming code.

In the second phase of screening 139 full-text manuscripts, we excluded 83 publications due to the eligibility criteria, not retrieval, or the language other than English.

The full-text reading phase included the remaining 56 publications in which we found 17 references and cited in this review. Those publications were part of specific events (e.g. [65]) presented at the 21st Annual meeting of the American Association for Artificial Intelligence), conferences not indexed in scientific databases for the given year (e.g. International Conference on Educational Data Mining in 2014) or works indexed using different keywords (e.g. [66]). Finally, this systematic review includes 73 publications and the original BKT publication.

To get a closer insight into the publications included in the review, we provided the yearly heatmap of the most frequent sources of BKT research in Table 2.4, in which 'Other' denotes sources that contributed to the review study with a single publication. The most common sources were scientific conferences, including the International Conference on Educational Data Mining (EDM), the International Conference of Intelligent Tutoring Systems (ITS), the International Conference of Artificial Intelligence in Education (AIED), the International Conference on User Modelling, Adaptation, and Personalization (UMAP), ACM Conference

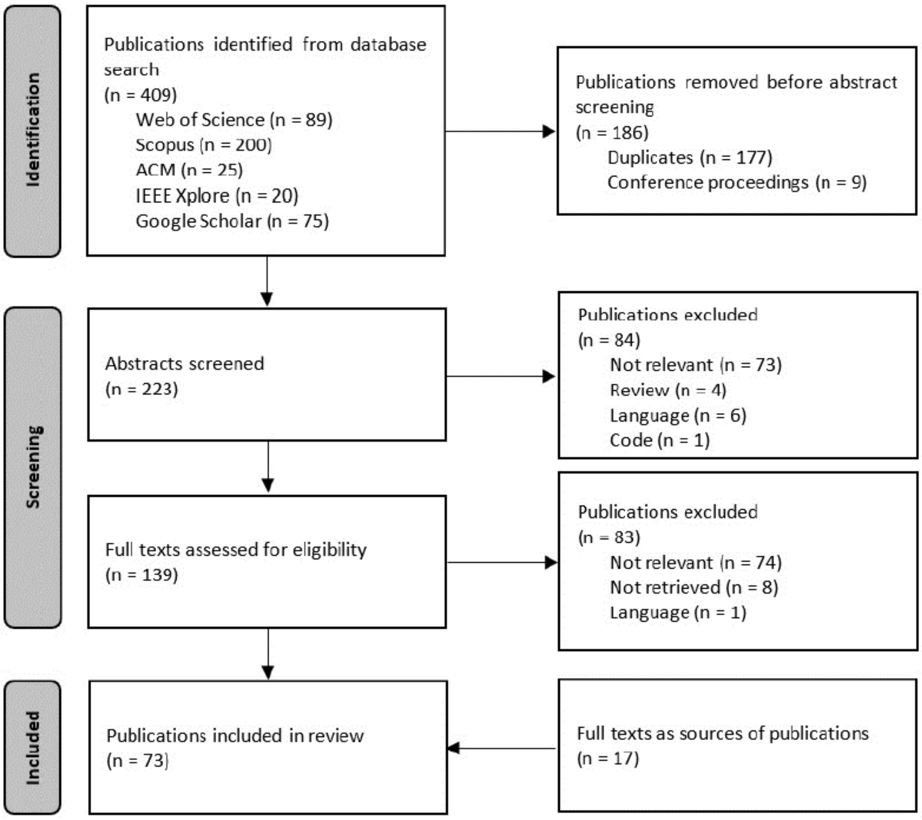

Figure 2.3. PRISMA flow diagram of the publication identification and screening process [1].

Table 2.4. Heatmap of the most frequent publication sources in the research of BKT enhancements [1].

Conf./ Journal	1995	2004	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2020	2021	2022	T
IC EDM					1	1	2	1	1	5	3	1	5		3				23
IC ITS		1			2				2		2		2						9
IC AIED				1						2				1			1	1	6
IC UMAP							2	1				1	2						6
ACM C L@S												2	1			1			4
IC LA&K												2	1						3
IEEE BigData														1	1				2
UMUAI	1				1														2
Other			1	1							3	3	1	2	3		2	2	18
Total (T)	1	1	1	2	4	1	4	2	3	7	8	9	12	4	7	1	3	3	73

on Learning at Scale (L@S), International Conference on Learning Analytics & Knowledge (LA&K), IEEE Conference on Big Data (IEEE BigData) and the User Modelling and User-Adapted Interaction (UMUAI) Journal. There was an increase in publications in 2008, 2010, and between 2013 and 2018.

To address RQ1 and elaborate on various enhancements of the BKT, we found the vanilla model assumptions to be appropriate review criteria. The vanilla assumptions derive from the architectural and educational context-based properties of the vanilla BKT model proposed by Corbett and Anderson in 1995 [9]. The architectural properties refer to the Hidden Markov model elements, including the nodes with corresponding states and the relationships between nodes (assumptions A01-A07 in the following text). The educational context-based properties include the vanilla assumptions on the knowledge component dependence, question difficulty and answer attempts (A08-A10).

The theory of knowledge inference in the vanilla model consists of the knowledge node with the binary learned and unlearned state (A01) and the performance node with the binary correct and incorrect state (A02). The prior knowledge, guessing, slipping, and learning parameters refer to expert-based probabilities estimated per skill (A03-A06). The model follows the no-forgetting paradigm by omitting the transition from learned to unlearned state (A07). The independent knowledge components (A08) refer to equally difficult questions used during the knowledge inference process (A09). Although a student may have multiple attempts to answer the question in the educational platform, the vanilla model counts only the first attempt (A10).

To address RQ1, we reviewed the computational methods used in the enhanced BKT models and architectural and educational context-based enhancements.

Regarding RQ2, each publication that proposed enhancements evaluated the approaches using datasets from specific educational platforms. Although the diversity and specificity of these studies did not allow a direct comparison of the reported results, this review study provides more insights into the evaluation approaches.

2.3.2 Results

This systematic overview of BKT enhancement aspects encompassed the identified research studies, which resulted in 62 enhanced BKT models. We noted some publications as multiple sources of the single enhanced BKT model (e.g. [67, 65]). For more than one source publication per model, we considered the year of the earlier publication as a model source year.

While some BKT models addressed the architectural and educational context-based properties of the vanilla BKT model (A01-A10), some enhancements extended its characteristics. Both types of enhancement aspects could also propose new computational methods. Therefore, we found it important to analyse architectural and educational context-based enhancements and computational methods separately.

The review of architectural and educational context-based properties, the overview of computational methods generally used for parameter estimation, and the evaluation approaches of enhanced BKT models are presented in the following subsections.

Table 2.5. Enhancement criteria used to review BKT models [1].

BKT Enhancement Aspects (EA)		**Vanilla BKT model Assumptions (A)**	
EA01	Knowledge states	A01	Knowledge component node in the Bayesian network includes the binary *learned* and *unlearned* state
EA02	Performance states	A02	Performance node in the Bayesian network includes the binary *correct* and *incorrect* state
EA03	Prior knowledge	A03	The prior knowledge probability is defined by experts and per skill
EA04	Guessing	A04	There is a probability of guessing defined by experts per skill
EA05	Slipping	A05	There is a probability of slipping defined by experts per skill
EA06	Learning	A06	The learning transition probability is defined by experts per skill
EA07	Forgetting paradigm	A07	The *no-forgetting paradigm* is followed meaning that there is no transition from a learned to an unlearned state
EA08	Domain knowledge properties	A08	Domain knowledge fractionates into *independent* knowledge components
EA09	Question difficulty	A09	Questions of each knowledge component are of *equal difficulty*
EA10	Multiple attempts	A10	The *first attempt* to answer the question counts during the modelling process
EA11	Student characteristics	Not included	
EA12	Tutor interventions	Not included	
EA13	Noise in data	Not included	

Table 2.6. The heatmap of the research on BKT enhancements [1].

BKT Enhancements (E)		**2004**	**2006**	**2008**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**T**
EA01	Knowledge states			1								1	2			1		5
EA02	Performance states			1	1					1	2					1		6
EA03	Prior knowledge				1			2	1	2	1	1						8
EA04	Guessing			1	1	1						1	1					5
EA05	Slipping			1	1	2						1	1					6
EA06	Learning							3					1					4
EA07	Forgetting paradigm		1	1		1				1	1		1					6
EA08	Domain know. prop.							1	3	1	3	1		1			2	12
EA09	Question difficulty					1		1	3	1	1	1	1					9
EA10	Multiple attempts			1	1			1	1	1	1				1			7
EA11	Student charact.			1	1		3		4	1	4		4			1	1	20
EA12	Tutor interventions		1	2	2					1	2	1	1					10
EA13	Noise in data	1								1							1	3
Total (T)		1	2	9	8	5	3	8	12	10	15	7	12	1	1	3	4	**101**

The architectural and context-based enhancements

To review the enhanced BKT models, we proposed the enhancement criteria in line with the vanilla BKT model assumptions. The enhancement criteria resulted from an iterative analysis of the identified research studies and represented a unique way of classifying the BKT enhancements. Besides those criteria found in the vanilla BKT model (enhancement aspects EA01-EA10 in Table 2.5), some enhancements extended the vanilla BKT model with new aspects. Additional vanilla BKT enhancement aspects included Student characteristics (EA11), Tutor interventions (EA12), and Noise in data (EA13). Table 2.5 shows the complete list of BKT enhancements and the related vanilla model assumptions.

Although each change in the Bayesian network architecture directly implied the update of BKT parameters, EA03-EA06 criteria encompassed BKT models focusing on the prior knowledge, guessing, slipping and learning BKT parameters, e.g. Contextual Guess and Slip method [68]. A yearly heatmap (Table 2.6) reviews BKT enhancements. The summarizations per year are presented in the last table row and per each enhancement in the last table column. Overall, fifty-four enhanced BKT models addressed 101 architectural and educational context-based properties of the vanilla BKT model.

The first enhanced BKT model emerged in 2004, a decade after the vanilla model. There

Table 2.7. The variations of investigated BKT enhancements [1].

#	BKT Enhancement Aspects (EA)	# Enhanced BKT models
1	EA11 Student characteristics	9
2	EA08 Domain knowledge properties	7
3	EA03 Prior knowledge	5
4	EA04 Guessing, EA05 Slipping, EA11 Student characteristics, EA12 Tutor interventions	3
5	EA09 Question difficulty, EA11 Student characteristics	2
6	EA11 Student characteristics, EA12 Tutor interventions	2
7	EA13 Noise in data	2
8-31	Other	24

was a decrease in the research after 2018, probably due to the COVID-19 pandemic. Much research work focused on enhancements to the vanilla model as Student characteristics (20 research studies), Domain knowledge properties (12 research studies), Tutor interventions (10 research studies), and Question difficulty (9 research studies). The most investigated enhancements extended the vanilla BKT model, precisely Student characteristics and Tutor interventions. The most investigated educational context-based enhancements referred to Domain knowledge properties and Question difficulty, while the most frequently investigated architectural enhancement included Prior knowledge.

Since each examined research study could enhance one or more of the proposed criteria, we analysed the most frequent variations of the investigated BKT enhancements. It is worth noting that a single criterion represents the most straightforward enhancement variation. The results are presented in Table 2.7, and variations found in a single research study are summarised as ‘Other’.

The results indicate that the single criteria of Student characteristics, Prior knowledge, and Domain knowledge properties were the most frequently investigated among 31 enhancement variations. The most frequent combination of enhancements found in 3 research studies included Guessing, Slipping, Student characteristics and Tutor intervention criteria. Table 2.8 shows the research related to each BKT Enhancement Aspect (EA).

Table 2.8. Enhanced BKT models per proposed Enhanced Aspect (EA) [1].

BKT EA	BKT models
EA01	Halpern et al., 2018 [69]; F. Liu et al., 2021 [70]; Schodde et al., 2017 [34]; Yudelson et al., 2008 [35]; Zhang & Yao, 2018 [2]
EA02	David et al., 2016 [71]; F. Liu et al., 2021 [70]; Ostrow et al., 2015 [72]; Y. Wang et al., 2010 [73]; Y. Wang & Heffernan, 2013 [74]; Z. Wang et al., 2016 [75]; Yudelson et al., 2008 [35]
EA03	Eagle, Corbett, Stamper, McLaren, Baker, et al., 2016 [76]; Eagle, Corbett, Stamper, McLaren, Wagner, et al., 2016 [77]; Eagle et al., 2017 [78]; Nedungadi & Remya, 2014 [79], 2015 [80]; Pardos & Heffernan, 2010 [81]; Song et al., 2015 [82]; S. Wang et al., 2017 [83]; Xu & Mostow, 2013 [84]; Yudelson et al., 2013 [3]
EA04	Agarwal et al., 2018 [85]; Baker et al., 2008a [66], 2008b [68], 2010 [30]; Pardos & Heffernan, 2011 [86]; Zhou et al., 2017 [87]
EA05	Agarwal et al., 2018 [85]; Baker et al., 2008a [66], 2008b [68], 2010 [30]; Pardos & Heffernan, 2011 [86]; Qiu et al., 2011 [88]; Zhou et al., 2017 [87]
EA06	Adjei et al., 2013 [89]; Baker et al., 2018 [90]; Sao Pedro et al., 2013 [91]; Yudelson et al., 2013 [3]
EA07	Beck et al., 2008 [67]; Chang et al., 2006 [65]; Halpern et al., 2018 [69]; Khajah et al., 2016 [63]; Nedungadi & Remya, 2015 [80]; Qiu et al., 2011 [88]; Yudelson et al., 2008 [35]
EA08	Chan et al., 2022 [92]; González-Brenes et al., 2014 [36]; Hawkins & Heffernan, 2014 [93]; Huang et al., 2016 [38]; Huang & Brusilovsky, 2016 [94]; Khajah et al., 2016 [63]; MacHardy, 2013 [95]; MacHardy & Pardos, 2015 [96]; Meng et al., 2019 [97]; Sao Pedro et al., 2013 [91]; Sun et al., 2022 [98]; Z. Wang et al., 2016 [75]
EA09	Baker et al., 2018 [90]; David et al., 2016 [71]; González-Brenes et al., 2014 [36]; Khajah, Huang, et al., 2014 [99]; Khajah, Wing, et al., 2014 [37]; Ostrow et al., 2015 [72]; Pardos et al., 2013 [100]; Pardos & Heffernan, 2011 [86]; Zhou et al., 2017 [87]
EA10	Bhatt et al., 2020 [101]; González-Brenes et al., 2014 [36]; Pardos et al., 2013 [100]; Yudelson et al., 2008 [35]
EA11	Agarwal et al., 2018 [85]; Baker et al., 2008a [66], 2008b [90], 2010 [30]; Corrigan et al., 2015 [102]; Eagle et al., 2018 [103]; Gorgun & Bulut, 2022 [104]; Halpern et al., 2018 [69]; Khajah et al., 2016 [63]; Khajah, Huang, et al., 2014 [99]; Khajah, Wing, et al., 2014 [37]; Lee & Brunskill, 2012 [105]; Lin et al., 2016 [32]; Lin & Chi, 2016 [33]; Nedungadi & Remya, 2014 [79]; Pardos et al., 2012 [106]; Rau & Pardos, 2016 [107]; Spaulding et al., 2016 [108]; Y. Wang & Heffernan, 2012 [109]; Xu et al., 2013 [84]; Yudelson et al., 2008 [35]
EA12	Agarwal et al., 2018 [85]; Baker et al., 2008a [66], 2008b [68], 2010 [30]; Beck et al., 2008 [67]; Chang et al., 2006 [65]; Lin et al., 2016 [32]; Lin & Chi, 2016 [33]; Ostrow et al., 2015 [72]; Rau & Pardos, 2016 [107]; Schodde et al., 2017 [34]; Y. Wang et al., 2010 [73]; Y. Wang & Heffernan, 2013 [74]; Yudelson et al., 2008 [35]
EA13	Beck & Sison, 2004 [31]; Falakmasir et al., 2015 [110]; Gorgun & Bulut, 2022 [104]

Computational methods

Computational methods used in the proposed BKT approaches generally referred to the estimation of BKT parameters. Some models did not have to interfere with the vanilla model assumptions but primarily addressed the computational challenges, e.g., the Dirichlet priors method [111, 112]. Overall, 56 enhanced BKT models reported the use of computational methods.

Table 2.9 presents the results of reviewing computational methods for researching BKT enhancements with over two applications.

Table 2.9. Computational methods used in the research of BKT enhancements [1].

#	Computational methods with over two applications	# Enhanced BKT models
1	Expectation-Maximization method	24
2	Markov Chain Monte Carlo (MCMC) method	5
3	Brute force method	4
4	K-means clustering	4
5	Contextual Guess and Slip method	3
6	Knowledge Heuristics and Empirical probabilities method	3
7-31	Other	32

Computational methods enhanced the skill-based estimations of BKT parameters used in the vanilla model. The Expectation-Maximization method, first used in 2006, practically became the standard (24 research studies). The other computational methods included the Monte Carlo method (5 research studies), the Brute force method (4 research studies), K-means clustering (4 research studies), the Contextual Guess and Slip method (3 research studies), and the Knowledge Heuristics with Empirical Probabilities method (3 research studies).

Evaluation approaches

Concerning the evaluation approaches, we reviewed educational platforms and performance measures used in researching BKT enhancements. Table 2.10 shows a yearly heatmap of the educational platforms used in the reviewed publications (Other denotes platforms with a single application).

Table 2.10. The datasets collected from educational platforms and used in the BKT research [1].

Educational platforms	2004	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	T
Assistments						2	2	2	1	3	3	1	1	1			1	2	19
Cognitive Tutor				1	1	1	2	1	2	2	2	2		1	1		2	1	19
Simulated data						1			2			1		2	1				7
MOOC									1		1	2		1					5
JavaGuide									1	1		1							3
Reading Tutor		1	1							1									3
Andes Tutor										2									2
Inq-ITS									1	1									2
Robot Tutor												1	1						2
Other	1			1						2	1	6	2	3	1	1			18
Total (T)	1	1	1	2	1	4	4	3	8	12	7	14	4	8	3	1	3	3	**80**

Besides ITSs, we found the application of the BKT model enhancements in Massive Open Online Courses (MOOCs), game-based platforms, and online learning platforms in human resources training. The research on the BKT enhancements typically included the Cognitive Tutor (19 research studies) and the ASSISTments (19 research studies). Other educational platforms with over two applications were MOOCs (5 research studies) and simulated datasets (7 research studies). The MOOC environments included edX, Coursera, Khan Academy and Junyi Academy.

As for the domain, the examined datasets related to Math (38 research studies), Language learning and Programming (per 7 research studies), Genetics, Physics and Engineering (per

Table 2.11. Performance measures used in the research of BKT enhancements [1].

#	Performance measures with over two applications	# Enhanced BKT models
1	RMSE	28
2	AUC	23
3	Accuracy	20
4	MAE	8
5	Correlation	3
6-14	Other	13

3 research studies), Science (per 2 research studies), and Medicine and Chemistry (per single research study).

Regarding the performance measures used, Table 2.11 shows the most frequent measures in the research of BKT enhancements with over two applications.

The most frequently used performance measures included the RMSE measure (Root Mean Square Error, 28 research studies), the AUC (Area Under Curve, Receiver Operating Characteristics curve, 23 research studies) and the Accuracy measure (20 research studies). These performance measures are frequently used metrics for classification tasks in the machine learning field.

3 METHODOLOGY

The thesis encompassed several tasks to investigate the proposed hypotheses (Figure 3.1). Initially, data was collected within an authentic educational setting (in situ) for each formative assessment, student and question. In addition to capturing the binary correctness of student answers, data on time spent on the tasks and the number of code evaluations was extracted. This process resulted in the Initial BKT report, which included Quiz ID, Student ID, Question ID, Student Performance, Time on Task (sec), Cumulative time (sec), Response evals and Cumulative evals.

Next, the Initial BKT report was used to classify student answers into multiple classes. The classification models complemented the report with features based on time on task (*Multiclasses_time*) and a combination of time on task and the number of code evaluations (*Multiclasses_time_evals*).

Subsequently, the student-level training subset from the previous output was used to fit the BKT parameters, including Prior, Guess, Slip, Learn, and Forget. The testing subset then evaluated the proposed BKT models, addressing Hypothesis 1 on the feasibility of the approach. Finally, a framework for ranking BKT models was introduced to improve the vanilla BKT model, addressing Hypothesis 2.

The results of the previously mentioned research tasks facilitated the BKT-based formative assessment in the widely used Modular Object-Oriented Dynamic Learning Environment (Moodle) LMS [113] through the development of the BKT Application Programming Interface (API) and the prototype modules.

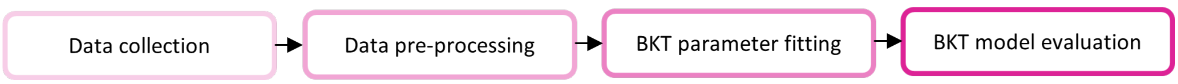

Figure 3.1. Research tasks conducted to investigate the hypotheses.

3.1 Data collection

Data collection in educational settings typically involves conducting in situ experiments that do not disrupt the natural behaviour of students and teachers. Educational interventions and phenomena are often evaluated using quasi-experimental designs that do not rely on random assignment. This empirical research employed an in situ experiment conducted

Table 3.1. The domain knowledge components used in the formative assessments.

# Topic	Domain Knowledge Component	# Subtopic(s)
1	DKC 01 Basic data types	
	DKC 02 Variables	
2	DKC 03 Operators and expressions - Arithmetic	1
	DKC 04 Operators and expressions - Relational	1
	DKC 05 Operators and expressions - Logic	1
	DKC 06 Operators and expressions - Other	1
	DKC 07 Operators and expressions - Priority	1
3	DKC 08 User input	1, 2
	DKC 09 User output	1, 2
4	DKC 10 Conditions (if-elif-else)	1, 2, 3
5	DKC 11 Loops (for) - Basic	1, 2, 3
	DKC 12 Loops (for)	1, 2, 3, 4
6	DKC 13 Loops (while) - Basic	1, 2, 3
	DKC 14 Loops (while)	1, 2, 3, 4
	DKC 15 Loops - Basic algorithms	1, 2, 3, 4, 5
7	DKC 16 Functions - Basic	1, 2, 3, 4, 5, 6
	DKC 17 Functions	
8	DKC 18 Lists - Basic 22/23	1, 2, 3, 4, 5, 6
	DKC 19 Lists - Basic	
	DKC 20 Sorting algorithm	
	DKC 21 Lists	1, 2, 3, 4, 5, 6, 7
9	DKC 22 Files - Basic 22/23	1, 2, 3, 4, 5, 6
	DKC 23 Files (including Functions)	1, 2, 3, 4, 5, 6, 7
	DKC 24 Files (including Lists)	1, 2, 3, 4, 5, 6, 7, 8
	DKC 25 Files (including Sorting algorithm)	

in the natural environment of a one-semester undergraduate course, "Introduction to Programming in Python". The course at the Faculty of Science, University of Split, enrolled 174 students, representing a non-random experimental sample. Formative assessment data collected through the open-source Moodle LMS (version 4.1.5+) [113] was used alongside student performance on standard summative evaluations, such as midterm and final exams. Students provided informed consent for their participation in the study, permitting the use of formative assessment data for research purposes.

The research investigated the granularity of the BKT Domain Knowledge Components (DKC) associated with that higher education course. A DKC was defined as a concept taught weekly, with a single week potentially encompassing multiple DKCs, each corresponding to a question pool that assessed the associated concept. Table 3.1 presents the DKCs examined in our experiment.

Following the weekly lectures and laboratory exercises, students were expected to master the knowledge individually. A day before the laboratory exercises, they had a single opportunity to assess their knowledge through a formative assessment designed for Self-Practice (SP). Then, a Controlled Environment (CE) formative assessment was conducted at the beginning of the laboratory exercises. SP and CE formative assessments together foster independent learning and self-reflection. While SP helps students build problem-solving skills and autonomy, CE assessments provide feedback that guides both teachers and students in addressing learning gaps, promoting continuous improvement. This research proposed en-

W 01-02	W 03	W 04	W 05	W 06	W 07	W 08	W 09	W 10	W 11	W 12	W 13	W 14	W 15
Informed consent	DKC 01-07 SP DKC 01-07 CE	DKC 08-10 SP DKC 08-10 CE	DKC 11-12 SP DKC 11-12 CE	DKC 13-14 SP DKC 13-14 CE	Midterm exam	DKC 15 SP DKC 15 CE	DKC 16-17 SP DKC 16-17 CE	DKC 18-19 SP DKC 18-19 CE	DKC 20-21 SP DKC 20-21 CE	DKC 22-23 SP DKC 22-23 CE	DKC 24-25 SP DKC 24-25 CE	Usability study	Final exam

Figure 3.2. Research protocol – course timeline.

Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu
Lab. exercises G1-G2			Lab. exercises G3-G5 Lectures G1-G5			DKC SP G1-G2	Lab. exercises G1-G2 DKC CE G1-G2		DKC SP G3-G5	Lab. exercises G3-G5 DKC CE G3-G5 Lectures G1-G5

Figure 3.3. Research protocol – DKC timeline.

hanced BKT modelling of student knowledge based on CE assessments. The collected SP data was discussed in the context of in situ experiment guidelines.

Each formative assessment comprised 20 questions, all of equal difficulty, and presented in a randomised order. The duration of the formative tests varied: DKCs 01-07 were allotted 10 minutes, DKCs 08-10 were given 15 minutes, and DKCs 11-25 had a duration of 20 minutes.

Figure 3.2 illustrates the research protocol for the course over 15 weeks (W). During the initial two weeks, students were asked to provide informed consent. Assessments were conducted for one or more DKCs taught in the preceding week. The course included typical summative assessments such as a midterm exam during the 7th week and a final exam at the end of the semester (15th week). Students participated in an anonymous usability study one week before the final exam.

Figure 3.3 outlines the weekly research protocol. In addition to two-hour lectures, students engaged in two-hour laboratory exercises, with the class divided into five groups (G1-G5). Each DKC was introduced during the regular lectures and corresponding laboratory exercises. The following week, students completed SP formative assessments followed by CE formative assessments.

3.1.1 Formative assessment using Virtual Programming Lab

The questions included in the formative assessments consisted of short programming tasks presented using Virtual Programming Lab (VPL) activity modules [114]. The core VPL (version 4.1.1) serves as an activity module type of Moodle extension, facilitating the management of programming tasks. In contrast, the VPL Question (version 1.7.0) is a specific

Figure 3.4. The VPL activity creation.

question type within Moodle that incorporates VPL-based tasks into quizzes. To facilitate student code execution and evaluation within our experimental environment, 6 VPL servers, each equipped with 2vCPUs and 4GB of RAM were utilized.

Since its emergence in 2012, the use of VPL modules has continuously grown worldwide, with over 2,000 sites reported in January 2023 [5]. Initial studies indicated that the module was both flexible and robust, enabling advanced methods for assessing student programming submissions [115, 116, 117]. Although automatic grading required considerable preparation time, the process required minimal effort once this phase was complete. Regarding interpreting increased code evaluations—whether they indicate deeper student engagement or disengaged behaviour (student clickers)—the VPL functionality that counted new evaluations only when students modified their code was found particularly useful. An example of configuring a VPL-based question within the Moodle LMS environment is presented in Figures 3.4- 3.8.

Initially, the teacher created a VPL activity to process questions and define the execution files (Figure 3.4). Run and evaluation options are set at this point, and the optional test scripts are added to test student programs. The VPL activities remain available for use by the VPL questions but are not visible to students in the Moodle course.

Then, a teacher creates a new VPL question in the question bank by configuring general information, question templates, answer templates, and settings for teacher correction and evaluation (Figures 3.5- 3.7). At this point, linking the question to the unique VPL activity is important to ensure the proper count of student code evaluations (Figure 3.5). Besides question text, the presented test cases reflect additional details to consider when answering a question. VPL Question Template refers to the linked unique VPL activity, and the ANSWER placeholder represents the student code to be executed. The Answer Template

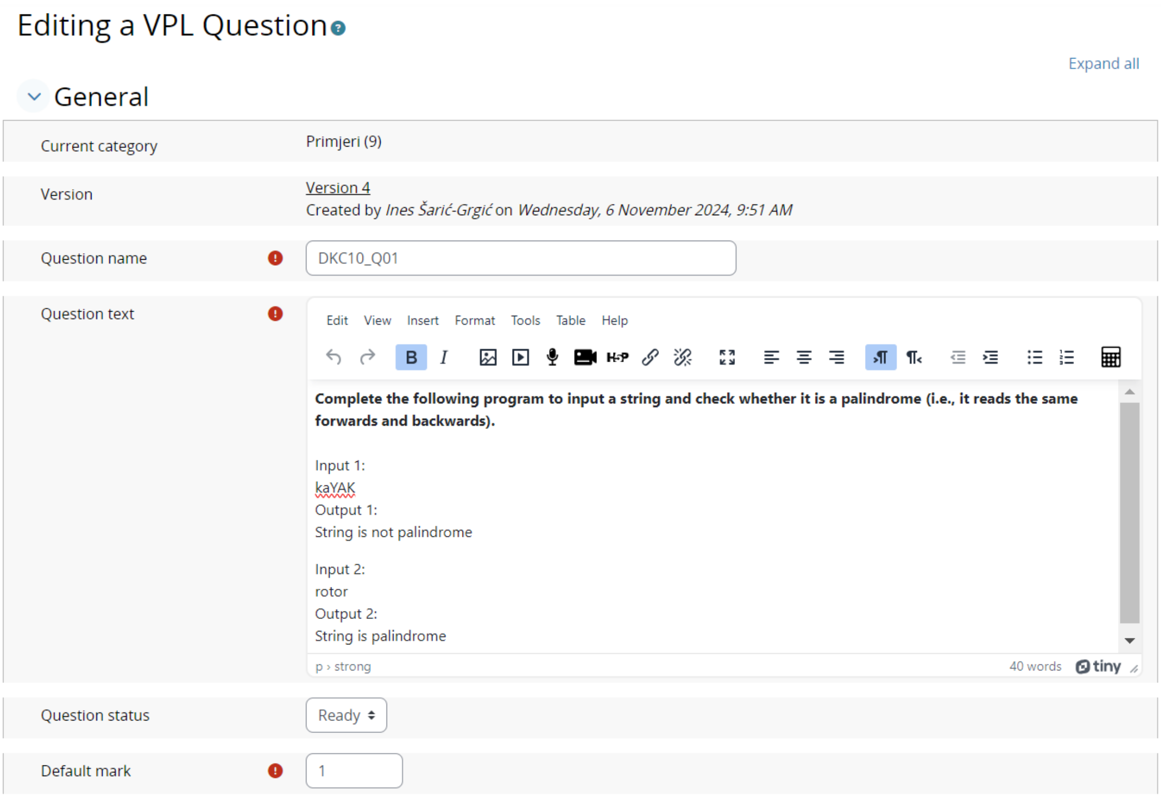

Figure 3.5. The VPL question creation – Question text.

represents the optional code presented to students, and the Teacher Correction presents the correct answer to be evaluated and presented to students as feedback. Figure 3.7 shows the test cases to be used during the evaluation.

From a student's point of view, VPL facilitated the modification and execution of the provided code and preliminary validation through pre-prepared test cases. While executing the code allowed for arbitrary input, the pre-check function assessed student programs using a set of predefined test cases (Figure 3.8).

In 21 out of 25 CE formative assessments, data was collected to fit BKT models. The four CE assessments (DKC 18 Lists - Basic 22/23, DKC 20 Sorting algorithm, DKC 22 Files - Basic 22/23, and DKC 24 Files including Lists) were used for testing the Moodle LMS prototype modules. From the student's perspective, the latter assessments were designed to demonstrate the idea of potential time efficiency in adaptive assessment. Following this introduction to adaptive assessment and their overall learning experience, students participated in the usability study before the final exam.

3.2 Data preprocessing

The data mining methodology in the context of data pre-processing, BKT parameter fitting and evaluation was implemented using a hosted Jupyter Notebook service (Google Colab)

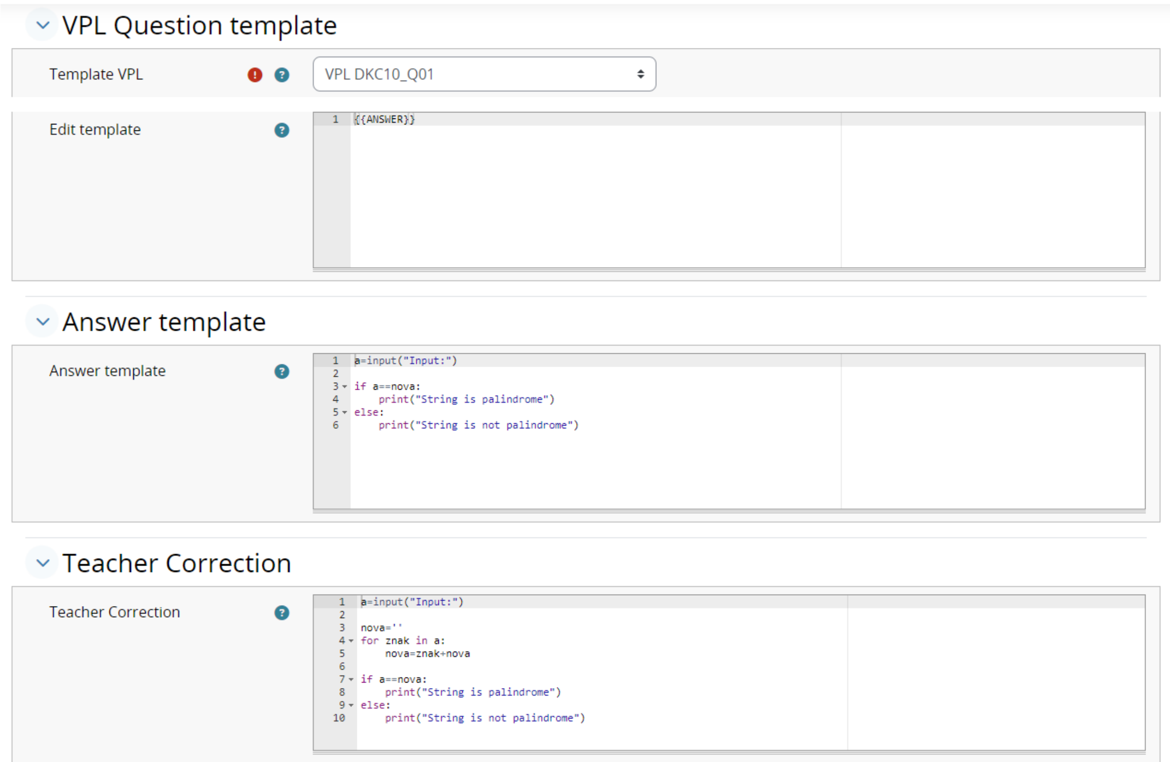

Figure 3.6. The VPL question creation – the VPL activity link (Template VPL), answer template and teacher correction.

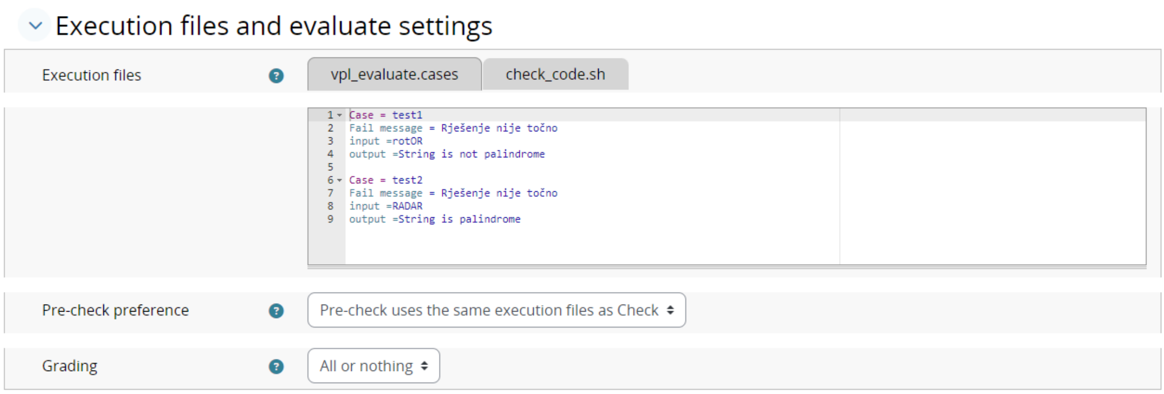

Figure 3.7. The VPL question creation – Execution and evaluation.

and Python libraries, including pandas (version 1.4.1), numpy (version 2.0.1), scikit-learn (version 1.5.1), scipy (version 1.14.0) and pyBKT (version 1.4.1).

In the data pre-processing phase, outliers were identified and removed based on the cumulative time spent on each task. The cumulative time spent on a task was defined as the duration from the start of the assessment to the moment a student completes a specific question. If a student deviated from the standard learning trajectory—such as skipping questions or using the browser's back button—the cumulative time might not have aligned with expected norms. Consequently, any assessment attempts exhibiting a cumulative time that

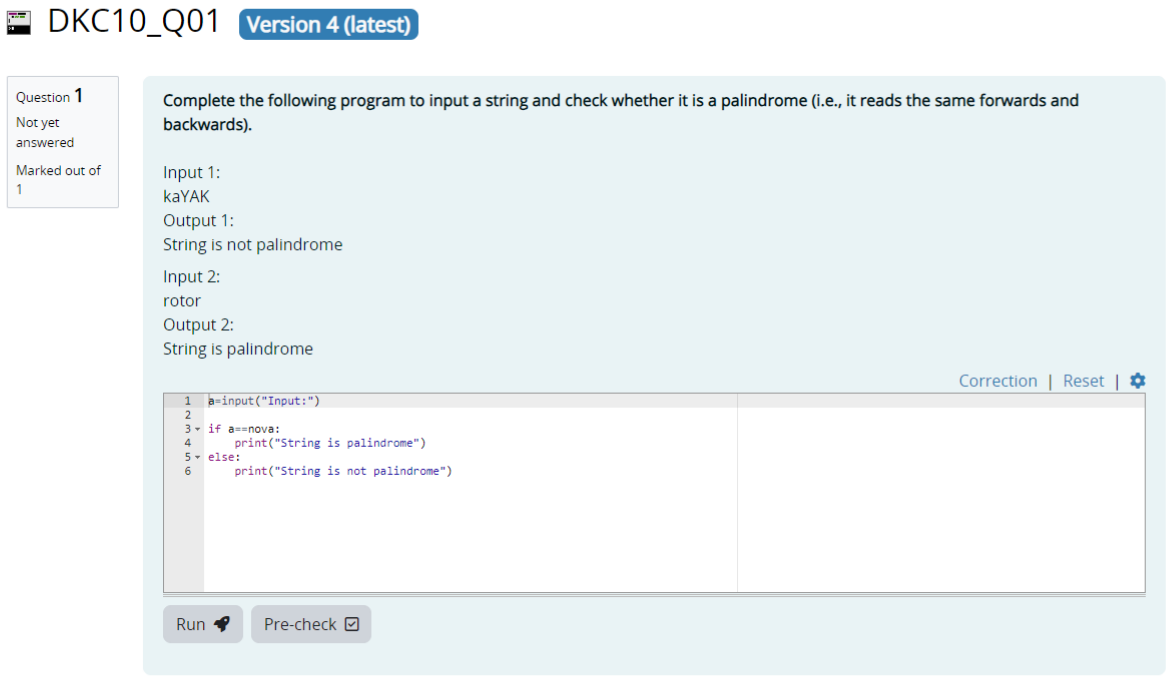

Figure 3.8. An example of the VPL-based "complete the given code" question.

significantly diverged from the total assessment duration, exceeding an arbitrary threshold of 10 seconds, were excluded from the dataset.

An answer opportunity refers to a single question in a formative assessment. During the pre-processing, instances where not all answer opportunities were completed by students were designated as incorrect answers. For these instances, the last recorded time spent on the task and the number of evaluation attempts were retained.

The pre-processed datasets were divided into training (70%) and testing (30%) subsets to investigate multiple classes of student answers. The classification of student answers relied on several key features: time spent on the task, cumulative time spent on the task, the number of student code evaluations, and cumulative code evaluations. Here, time spent on the task denoted the total seconds allocated to each answer opportunity (question), while the number of code evaluations referred to the number of code checks using pre-defined test cases. The cumulative values of these features represented the accumulated time and number of evaluations from the start of the assessment up to the current question.

The Decision Tree method was employed for student answer classification due to its high interpretability and widespread application. The default Decision Tree method was based on the CART (Classification and Regression Trees) algorithm from the scikit-learn Python library. Information gain and entropy were utilized as the splitting criterion measures. Based on experimental results indicating the minimal depth required for all examined DKCs to achieve multiple classes, a maximum tree depth of 5 was established. Standard performance metrics—AUC, F1, Precision, Recall, and Accuracy—assessed classification model outcomes.

Moreover, to validate and evaluate the enhanced BKT models, the pre-processed datasets were divided into student-level training (70%) and testing (30%) subsets. The student data was shuffled to ensure that different students appear in the training and testing sets randomly. However, individual answer opportunities were kept intact and placed entirely in either the training or testing set based on the subset they belong to.

The BKT models were fitted and cross-validated using the student-level training subset, while their performance was assessed using the unseen data from the testing subset.

3.3 BKT parameter fitting

Cognitive modelling creates computational simulations of human thought processes to replicate how the brain handles information, decision-making, and problem-solving. These models provide insights into human behaviour and improve predictions of cognitive functions. The pyBKT is an open-source, accessible and computationally efficient Python library for cognitive modelling [62]. It encompasses the BKT model and its potential variants, providing a means of estimating students' cognitive mastery from sequences of problem-solving activities. Experiments conducted with the library examined the accuracy of synthetic model fitting. The results indicated that increasing the number of students had a greater impact on reducing fit error than extending the length of the task sequences assigned to students. A sample size of 50 was found to be sufficient for achieving convergence to canonical parameter values, regardless of the average sequence length per student. Additionally, a sequence length of 15 was found sufficient for minimizing inaccuracies in worst-case mastery estimation [62].

PyBKT was utilized to fit, cross-validate and evaluate proposed BKT models. It incorporated the widely used Expectation-Maximization (EM) method for model fitting. Although this method has practically become the standard in the field, increased research interest has highlighted potential challenges. Doroudi and Brunskill [118] identified that the BKT model was susceptible to the local optima problem under mild conditions on the parameters. Additionally, it was revealed that the EM method was prone to semantic model degeneracy, which could render the BKT model inconsistent with its underlying conceptual assumptions [113]. Model degeneracy was characterized by Guess and Slip parameter probabilities exceeding 0.5. The literature suggests limiting the parameter space by bounding these parameters to a maximum of 0.5 to reduce degeneracy caused by excessive answer opportunities, often interpreted as under- or over-practice.

Also, using multiple EM algorithm iterations with different initialisations helps to ensure more robust and stable BKT parameters. Setting numfits pyBKT parameter to a higher value provides multiple EM initializations enabling the best-fitting parameters less influenced by the randomness inherent in the EM algorithm.

Using the training data subset, the 18 BKT models were fit for adaptive assessments (Ta-

Table 3.2. The BKT models for adaptive assessment.

BKT model	pyBKT parameters (# parameters to model)	
#01 vanilla	Prior, Guess, Slip, Learn, Forgets=0	4
#02 vanilla+forgets	Prior, Guess, Slip, Learn, Forgets	5
#03 multigs T	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn, Forgets=0	6
#04 multigs+forgets T	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn, Forgets	7
#05 multilearn T	Prior, Guess, Slip, Learn_Class0, Learn_Class1, Forgets_Class0=0, Forgets_Class1=0	5
#06 multilearn+forgets T	Prior, Guess, Slip, Learn_Class0, Learn_Class1, Forgets_Class0, Forgets_Class1	7
#07 multigs+multilearn T	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn_Class0, Learn_Class1, Forgets_Class0=0, Forgets_Class1=0	7
#08 multigs+multilearn+forgets T	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn_Class0, Learn_Class1, Forgets_Class0, Forgets_Class1	9
#09 multiprior T	Prior=0, Guess, Slip, Learn_Class1, Learn_Class2, Learn_Default, Forgets_Class1=0, Forgets_Class2=0, Forgets_Default=0	5
#10 multiprior+forgets T	Prior=0, Guess, Slip, Learn_Class1, Learn_Class2, Learn_Default, Forgets_Class1, Forgets_Class2, Forgets_Default	8
#11 multigs TE	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn, Forgets=0	6
#12 multigs+forgets TE	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn, Forgets	7
#13 multilearn TE	Prior, Guess, Slip, Learn_Class0, Learn_Class1, Forgets_Class0=0, Forgets_Class1=0	5
#14 multilearn+forgets TE	Prior, Guess, Slip, Learn_Class0, Learn_Class1, Forgets_Class0, Forgets_Class1	7
#15 multigs+multilearn TE	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn_Class0, Learn_Class1, Forgets_Class0=0, Forgets_Class1=0	7
#16 multigs+multilearn+forgets TE	Prior, Guess_Class0, Guess_Class1, Slip_Class0, Slip_Class1, Learn_Class0, Learn_Class1, Forgets_Class0, Forgets_Class1	9
#17 multiprior TE	Prior=0, Guess, Slip, Learn_Class1, Learn_Class2, Learn_Default, Forgets_Class1=0, Forgets_Class2=0, Forgets_Default=0	5
#18 multiprior+forgets TE	Prior=0, Guess, Slip, Learn_Class1, Learn_Class2, Learn_Default, Forgets_Class1, Forgets_Class2, Forgets_Default	8

ble 3.2). While the time aspect could be applied to any question type, such as multiple-choice or essay questions, the number of code evaluations was specific to the Python programming. Table 3.2 presents the originally defined pyBKT parameters for each BKT model. When the forgetting mechanism was not enabled, the Forgets parameter was set to zero. The multigs and multilearn models utilised Class0 and Class1 parameter probabilities. In this research, 'multi classes' refers to an assessment environment centred on the binary correctness of student answers. In the multiprior models, the Prior parameter probabilities were set to zero, while different Learn probabilities were defined as Class1, Class2 and Default. Additionally, Class1 and Class2 represented parameter probabilities related to the binary correctness at the first answer opportunity, whereas the Default class probabilities applied to subsequent answer opportunities.

The vanilla model is the baseline BKT model, referred to as #01 vanilla in Table 3.2. The first enhancement included the addition of the forgetting parameter (#02 vanilla+forgets). Both models considered a single parameter class based on the correctness of the answer to model knowledge mastery.

Two groups of BKT models were examined based on the features they considered when modelling knowledge mastery. In addition to answer correctness, the first group of models (#03 - #10) incorporated the features of time spent on task and cumulative time. These time-feature-based models include T tag in the model names. The #03 multigs T model included different Guess and Slip parameter probabilities (Class0 and Class1), while its enhanced version incorporated forgetting (#04 multigs+forgets T). The #05 multilearn T model introduced different Learn parameter probabilities (Class0 and Class1), with its enhanced version allowing multi class forgetting (#06 multilearn+forgets T). The #07 multigs+multilearn T model combined multi class Guess, Slip, and Learn parameter probabilities (Class0 and Class1), and its enhanced version included multi class forgetting (#08 multigs+multilearn+forgets T). The #09 multiprior T model defined multi-class Learn parameter probabilities for the first answer opportunity (Class1 and Class2), interpreting it as a different level of student pre-knowledge (Prior). The multiprior model also had an enhanced forgetting version (#10 multiprior+forgets T).

The second group of enhanced BKT models (#11 - #18) further considered the number and cumulative number of code evaluations during classification. The time- and evaluation-feature-based models include TE tags in the model names. Similar to the previous group, models #11 multigs TE, #12 multigs+forgets TE, #13 multilearn TE, #14 multilearn+forgets TE, #15 multigs+multilearn TE, #16 multigs+multilearn+forgets TE, #17 multiprior TE and #18 multiprior+forgets TE were proposed.

3.4 BKT model evaluation

Finally, the performance of BKT models was assessed in terms of their ability to predict student performance and estimate knowledge mastery. The power of the BKT models to predict student performance was cross-validated using the training subset and evaluated with the testing subset. The ability to estimate knowledge mastery was analysed using the complete and the testing subsets. The complete dataset was also used to examine the average and ideal number of answer opportunities (model convergence).

Standard metrics, including RMSE, AUC, and Accuracy, were used to evaluate predictive power. The 5-fold cross-validation combined the fitting and evaluation of BKT models, while the evaluation encompassed unseen data. The nonparametric Wilcoxon Signed Rank Test was also applied to determine statistically significant differences between the baseline vanilla BKT model and each BKT model across DKCs.

The relationship between adaptive BKT probabilities and student performance in formative and summative assessments was analysed for knowledge mastery estimation. Formative assessments referred to complete student answer sequences, while summative assessments referred to student performance on typical midterm and final exams. These relationships were measured using a BKT mastery threshold of 0.95, correlation measure, corresponding p-values of statistical significance, and the F1 score.

The framework for ranking the BKT model relied on composite scores, which were based on the model's capacity to predict student performance and estimate knowledge mastery in a timely manner.

The composite scores included several research metrics presented in the study. RMSE and AUC metrics were used for prediction, whereas correlation and F1 scores were employed for formative and summative assessments to assess knowledge mastery. In evaluating model performance, the BKT ranking incorporated information on model convergence based on the learning path, represented as the sequence of binary correctness in student answers within the assessment. Model convergence was measured by the average number of answer opportunities across the dataset, as well as the ideal number of answer opportunities needed to reach knowledge mastery. The ideal learning path, in this context, reflects a scenario in which a student consistently responds correctly within a motivational framework aligned with the BKT parameter probabilities.

The aforementioned metrics were normalised using min-max scaling, and for metrics where lower values are preferable (such as RMSE and answer opportunity metrics), the values were inverted. Two types of composite scores were reported, including balanced weights. The first composite score considered data from formative and summative assessments (midterm and final exams), while the second score was based solely on formative assessment data.

3.5 Methodological instruments

The first hypothesis posited the feasibility of tracing student knowledge through Bayesian modelling by leveraging the features of time spent on the task and the number of code evaluations. The feasible BKT model is successfully fit and evaluated by considering each research task's requirements and potential limitations. In the Data collection task, within specific experimental settings, the features of time spent on the task and the number of code evaluations are successfully extracted. The Data pre-processing task ensures the utilisation of appropriate data and the classification of student answers. Subsequent BKT parameter fitting and model evaluation tasks confirm the feasibility of the BKT model.

Moreover, the extended BKT modelling enhancing the vanilla BKT model was observed to address the second hypothesis. A framework for BKT model ranking based on the research results was proposed.

The feasible BKT model that overperforms the vanilla model per each DKC takes into account the following research results:

- regarding the prediction of student performance: lower and statistically significant RMSE across DKCs; higher and statistically significant AUC across DKCs
- for the estimation of knowledge mastery: higher and statistically significant Pearson correlation; higher F1 score
- student learning paths: lower average number of answer opportunities; lower number of answer opportunities in the ideal student learning path.

The following Results and discussion section systematically addresses each of the proposed hypotheses, presenting findings in a structured manner across successive subsections.

4 RESULTS AND DISCUSSION

This section presents the research findings, progressively addressing the proposed hypotheses, including the feasibility of tracing student knowledge using enhanced Bayesian modelling and improving the vanilla BKT model performance results. Sections 4.1 through 4.7 provide the data and analysis required to examine the hypotheses. Subsection 4.1 outlines the data collection, followed by the classification of student answers in subsection 4.2. Subsection 4.3 details the fitted BKT model parameters for each DKC, with Subsection 4.4 presenting student performance prediction results, subsection 4.5 covering knowledge mastery estimation and subsection 4.6 focusing on student learning path outcomes. Subsection 4.7 presents the BKT model ranking results. Discussion on the hypotheses is presented in Subsection 4.8.

4.1 Formative assessment data

Formative assessment data for Self-Practice (SP) and the Controlled Environment (CE) were analysed for each DKC using descriptive statistics, including central tendency, dispersion and reliability measures. The mean, Standard Deviation (SD), and median were used to represent central tendency and dispersion, while reliability was assessed through the Coefficient of Internal Consistency (CIC), Error Ratio (ER), and Standard Error (SE). Table 4.1 provides an overview of the dataset, including the number of students, the mean and standard deviation of percentage scores (0-100%), the median percentage score, and the coefficient of internal consistency, error ratio, and standard error measures.

Cronbach's alpha (CIC) was used to assess the consistency of test questions in evaluating the same concept (DKC). The results are presented on a percentage scale, with values above 75% considered satisfactory. Since each test was designed around a specific DKC, the high average coefficient of internal consistency value of 93.09% confirmed the reliability of this approach.

The error ratio, derived from a coefficient of internal consistency, indicates the proportion of the SD likely attributable to random factors rather than differences in knowledge mastery. The error ratio values exceeding 50% are considered unsatisfactory. The average error ratio of 25.47%, with a maximum of 42.64% for DKC 05 (Logic operators & expressions SP), suggested satisfactory results.

Table 4.1. Formative test statistics – the central tendency, dispersion and reliability measures.

Domain Knowledge Component	# Students	Mean %	SD %	Median %	CIC %	ER %	SE %
DKC 01 Basic data types SP	152	16.74	22.80	5.00	91.34	29.43	6.71
DKC 01 Basic data types CE	107	37.52	32.22	30.00	93.94	24.62	7.93
DKC 02 Variables SP	145	32.10	26.19	30.00	89.82	31.91	8.36
DKC 02 Variables CE	105	51.81	31.48	50.00	92.89	26.67	8.40
DKC 03 Arithmetic oper. & expr. SP	141	17.80	17.27	15.00	82.01	42.41	7.33
DKC 03 Arithmetic oper. & expr. CE	92	31.30	22.87	30.00	87.24	35.73	8.17
DKC 04 Relational oper. & expr. SP	138	26.85	32.49	7.50	95.75	20.62	6.70
DKC 04 Relational oper. & expr. CE	80	44.75	36.48	50.00	95.91	20.21	7.37
DKC 05 Logic oper. & expr. SP	136	11.62	14.78	5.00	81.82	42.64	6.30
DKC 05 Logic oper. & expr. CE	79	18.42	17.59	15.00	83.77	40.29	7.09
DKC 06 Other oper. & expr. SP	135	25.33	29.37	15.00	94.25	23.98	7.04
DKC 06 Other oper. & expr. CE	77	29.87	33.55	10.00	95.92	20.19	6.78
DKC 07 Priority of oper. & expr. SP	132	9.51	21.05	0.00	95.10	22.14	4.66
DKC 07 Priority of oper. & expr. CE	73	8.97	20.00	0.00	94.60	23.23	4.65
DKC 08 User input SP	137	37.30	28.28	35.00	91.00	30.00	8.48
DKC 08 User input CE	139	60.54	30.88	70.00	93.06	26.35	8.14
DKC 09 User output SP	134	13.40	18.46	5.00	88.29	34.22	6.32
DKC 09 User output CE	139	22.37	23.13	15.00	89.62	32.21	7.45
DKC 10 Conditions SP	133	22.48	24.34	15.00	91.44	29.26	7.12
DKC 10 Conditions CE	138	33.70	25.80	35.00	90.46	30.89	7.97
DKC 11 Loops (for) – Basic SP	131	44.62	35.00	45.00	95.24	21.82	7.64
DKC 11 Loops (for) – Basic CE	127	62.24	33.28	70.00	94.95	22.46	7.48
DKC 12 Loops (for) SP	129	20.93	30.13	0.00	95.80	20.50	6.18
DKC 12 Loops (for) CE	127	29.37	33.44	15.00	95.76	20.59	6.89
DKC 13 Loops (while) – Basic SP	128	20.43	24.39	10.00	91.44	29.26	7.14
DKC 13 Loops (while) – Basic CE	125	35.52	31.90	25.00	94.12	24.24	7.73
DKC 14 Loops (while) SP	126	14.37	22.32	5.00	92.43	27.51	6.14
DKC 14 Loops (while) CE	124	30.56	30.57	20.00	94.03	24.43	7.47
DKC 15 Loops – Basic algorithms SP	108	31.11	31.11	22.50	94.10	24.28	7.56
DKC 15 Loops – Basic algorithms CE	115	43.30	33.20	40.00	94.15	24.18	8.03
DKC 16 Functions – Basic SP	109	31.79	34.74	15.00	96.01	19.98	6.94
DKC 16 Functions – Basic CE	121	54.17	39.16	60.00	96.91	17.59	6.89
DKC 17 Functions SP	107	10.98	21.45	0.00	94.32	23.83	5.11
DKC 17 Functions CE	121	19.63	28.10	10.00	95.13	22.07	6.20
DKC 19 Lists – Basic SP	115	38.57	36.73	30.00	96.31	19.20	7.05
DKC 19 Lists – Basic CE	107	64.44	34.31	80.00	95.35	21.56	7.40
DKC 21 Lists SP	111	14.14	23.53	0.00	93.83	24.85	5.85
DKC 21 Lists CE	120	24.63	30.67	10.00	95.07	22.21	6.81
DKC 23 Files (including Functions) SP	111	27.75	37.63	0.00	97.78	14.90	5.61
DKC 23 Files (including Functions) CE	113	55.66	41.23	70.00	97.73	15.07	6.21
DKC 25 Files (including Sorting alg.) SP	92	24.08	30.64	5.00	95.56	21.06	6.45
DKC 25 Files (including Sorting alg.) CE	95	46.95	35.58	50.00	95.47	21.28	7.57
Mean	118.43	30.90	28.77	24.29	93.09	25.47	6.98
SD	19.44	14.85	6.44	22.25	3.78	6.50	0.93

The standard error was used to determine the range within which a student's score would likely fall on repeated attempts of a similar test. Lower standard error values indicate more precise tests, though reducing standard error below 5- 6% is challenging. A standard error of 8% corresponds to roughly a half-grade difference, with higher values indicating a risk of misgrading [119]. The results were satisfactory, with an average standard error of 6.98% and a maximum of 8.40% for DKC 02 (Variables CE).

Furthermore, data distribution was analysed using the Kolmogorov-Smirnov (KS) test for normality, as well as skewness (SDS) and kurtosis (SDK) (Table 4.2). The Kolmogorov-Smirnov test statistic (D) was calculated using an online tool [120], while Moodle LMS provided the skewness and kurtosis. Table 15 categorized data distribution as either Normally distributed (ND) or Not Normally Distributed (NND).

The Kolmogorov-Smirnov D statistic measures the degree of deviation of the sample distribution from a normal distribution. A higher D value indicates a lower likelihood that the data follows a normal distribution. The associated p-value quantifies this likelihood, where a low p-value suggests that the sample deviates from normality beyond what could be expected by chance.

The skewness and kurtosis provide further insights into the distribution's shape. The skewness assesses asymmetry, with negative values indicating a greater frequency of larger values and positive values reflecting a prevalence of smaller values. A general rule is that a skewness value beyond the range -2 to 2 indicates significant non-normality. The kurtosis measures the thickness of distribution tails, with positive values indicating a more peaked distribution and negative values suggesting a flatter distribution. As with skewness, kurtosis values outside the range -2 to 2 are considered abnormal, with values above 2 suggesting a distribution that is too peaked and values below -2 suggesting a distribution that is too flat.

Normal distribution was observed in 6 CE tests, including those on DKC 01 (Basic data types CE), DKC 02 (Variables CE), DKC 03 (Arithmetic operators & expressions CE), DKC 04 (Relational operators & expressions CE), DKC 10 (Conditions CE), and DKC 15 (Loops – Basic algorithms CE). Among these, five tests exhibited positive skewness and negative kurtosis, while the DKC 02 (Variables CE) had negative skewness.

Out of the 21 CE tests analysed, 15 deviated from the assumption of normal distribution (71.42%). These deviations included both positive skewness and negative kurtosis values, indicating a higher frequency of low test scores and a distribution characterized by a long tail without pronounced peaks. The marked non-normality was particularly evident in the DKC 07 (Priority of operators & expressions CE). In Table 4.2, the skewness and kurtosis values exceeding the thresholds of -2 and 2 were underlined, while instances of untypical negative skewness and positive kurtosis values were highlighted using italics.

The experimental data is publicly available at [121]. Furthermore, CE assessments of student knowledge served as the input for BKT-based cognitive modelling.

Table 4.2. Formative test statistics – data distribution measures.

Domain Knowledge Component	# Students	KS D	p-value	SDS	SDK
DKC 01 Basic data types SP	152	0.23058	< 0.00001 NND	1.6098	*2.1340*
DKC 01 Basic data types CE	107	0.12819	0.0542 ND	0.4811	-1.0942
DKC 02 Variables SP	145	0.11832	0.03166 NND	0.5903	-0.4777
DKC 02 Variables CE	105	0.10217	0.20832 ND	*-0.1120*	-1.2768
DKC 03 Arithmetic oper. & expr. SP	141	0.1505	0.00297 NND	0.8837	*0.2588*
DKC 03 Arithmetic oper. & expr. CE	92	0.08935	0.42984 ND	0.4039	-0.6622
DKC 04 Relational oper. & expr. SP	138	0. 25014	< 0.00001 NND	0.8697	-0.6384
DKC 04 Relational oper. & expr. CE	80	0.14147	0.07361 ND	0.0791	-1.5171
DKC 05 Logic oper. & expr. SP	136	0.26165	< 0.00001 NND	1.3929	*1.0127*
DKC 05 Logic oper. & expr. CE	79	0.18376	0.00829 NND	0.8830	*0.5309*
DKC 06 Other oper. & expr. SP	135	0.21573	< 0.00001 NND	0.9086	-0.3685
DKC 06 Other oper. & expr. CE	77	0.23974	0.00022 NND	0.7249	-0.8328
DKC 07 Priority of oper. & expr. SP	132	0. 38055	< 0.00001 NND	2.5264	*5.5827*
DKC 07 Priority of oper. & expr. CE	73	0.4003	< 0.00001 NND	2.6885	*7.2342*
DKC 08 User input SP	137	0. 13293	0.01428 NND	0.3343	-1.0796
DKC 08 User input CE	139	0.14841	0.00388 NND	*-0.4024*	-1.1124
DKC 09 User output SP	134	0.23567	< 0.00001 NND	1.8892	*3.6751*
DKC 09 User output CE	139	0.19386	0.00005 NND	1.2665	*1.1325*
DKC 10 Conditions SP	133	0. 17698	0.00041 NND	1.1570	*1.1233*
DKC 10 Conditions CE	138	0.10791	0.07458 ND	0.3393	-0.5613
DKC 11 Loops (for) – Basic SP	131	0.12397	0.0326 NND	0.1230	-1.3820
DKC 11 Loops (for) – Basic CE	127	0.13695	0.01539 NND	*-0.5853*	-0.9383
DKC 12 Loops (for) SP	129	0.28358	< 0.00001 NND	1.3054	*0.4287*
DKC 12 Loops (for) CE	127	0.18895	0.00019 NND	0.9023	-0.4955
DKC 13 Loops (while) – Basic SP	128	0.2002	0.00006 NND	1.3661	*1.1091*
DKC 13 Loops (while) – Basic CE	125	0.15296	0.00509 NND	0.7039	-0.6802
DKC 14 Loops (while) SP	126	0.27623	< 0.00001 NND	1.8365	2.7440
DKC 14 Loops (while) CE	124	0.16317	0.00236 NND	0.8681	-0.3958
DKC 15 Loops – Basic algorithms SP	108	0.15753	0.00829 NND	0.6774	-0.7662
DKC 15 Loops – Basic algorithms CE	115	0.12143	0.06178 ND	0.1355	-1.3294
DKC 16 Functions – Basic SP	109	0.23942	< 0.00001 NND	0.5506	-1.2554
DKC 16 Functions – Basic CE	121	0.17377	0.00115 NND	*-0.2265*	-1.5824
DKC 17 Functions SP	107	0.33929	< 0.00001 NND	2.6475	*7.0302*
DKC 17 Functions CE	121	0.24153	< 0.00001 NND	1.7970	*2.2656*
DKC 19 Lists – Basic SP	115	0.21742	0.00003 NND	0.5033	-1.4027
DKC 19 Lists – Basic CE	107	0.18968	0.00076 NND	*-0.5421*	-1.1978
DKC 21 Lists SP	111	0.31855	< 0.00001 NND	1.7391	*2.1057*
DKC 21 Lists CE	120	0.21733	0.00002 NND	1.1002	-0.0796
DKC 23 Files (including Functions) SP	111	0.33816	< 0.00001 NND	0.8716	-0.9737
DKC 23 Files (including Functions) CE	113	0.18739	0.0006 NND	*-0.3069*	-1.6070
DKC 25 Files (including Sorting alg.) SP	92	0.33215	< 0.00001 NND	1.1751	-0.0617
DKC 25 Files (including Sorting alg.) CE	95	0.17804	0.00417 NND	0.0134	-1.5326

4.2 Classification of student answers based on time on task and the number of code evaluations

During the Data pre-processing, outliers were addressed by examining the cumulative time spent on each task. Unanswered questions were categorized as incorrect to account for

Table 4.3. The number of students, outliers and not-reached answer opportunities per each DKC.

Domain Knowledge Component	# Students	# Students without outliers	Not-reached AO
DKC 01 Basic data types CE	107	99	259/1980 (13%)
DKC 02 Variables CE	105	104	133/2080 (6%)
DKC 03 Arithmetic oper. & expr. CE	92	87	238/1740 (13%)
DKC 04 Relational oper. & expr. CE	80	77	67/1540 (4%)
DKC 05 Logic oper. & expr. CE	79	73	145/1460 (9%)
DKC 06 Other oper. & expr. CE	77	76	110/1520 (7%)
DKC 07 Priority of oper. & expr. CE	73	71	224/1420 (15%)
DKC 08 User input CE	139	137	90/2740 (3%)
DKC 09 User output CE	139	129	534/2580 (20%)
DKC 10 Conditions CE	138	138	354/2760 (12%)
DKC 11 Loops (for) – Basic CE	127	127	71/2540 (3%)
DKC 12 Loops (for) CE	127	127	252/2540 (10%)
DKC 13 Loops (while) – Basic CE	125	120	333/2400 (13%)
DKC 14 Loops (while) CE	124	122	210/2440 (8%)
DKC 15 Loops – Basic algorithms CE	115	115	214/2300 (9%)
DKC 16 Functions – Basic CE	121	121	123/2420 (5%)
DKC 17 Functions CE	121	121	342/2420 (14%)
DKC 19 Lists – Basic CE	107	107	90/2140 (4%)
DKC 21 Lists CE	120	112	336/2240 (15%)
DKC 23 Files (including Functions) CE	113	113	145/2260 (6%)
DKC 25 Files (including Sorting alg.) CE	95	95	138/1900 (7%)

missed opportunities. Table 4.3 summarises the final number of students across each DKC.

In addition to student performance, features such as time on task, cumulative time on task, the number of code evaluations and the cumulative number of code evaluations were extracted for each question in the DKC test. The cumulative values represented the time interval up to the observed question (answer opportunity). It was expected that multi-class classification based on these additional features would further refine the binary evaluation of a student's answer accuracy.

The descriptive statistics for the time on task and cumulative time on task features for each DKC are presented in Table 4.4 and Figures 4.1- 4.2, including the mean, standard deviation, median, and maximum values for each feature.

Table 4.4. Descriptive test statistics related to the time on task and cumulative time on task features.

Domain Knowledge Component	Time on task				Cumulative time on task			
	Mean	SD	Median	Maximum	Mean	SD	Median	Maximum
DKC 01 Basic data types CE	25.18	24.53	18.00	269.00	316.93	171.78	321.00	601.00
DKC 02 Variables CE	22.96	21.49	18.00	317.00	269.41	158.71	261.50	600.00
DKC 03 Arithmetic oper. & exp. CE	25.72	25.03	19.00	197.00	332.05	184.27	343.50	601.00
DKC 04 Relational oper. & exp. CE	19.88	18.73	15.00	202.00	236.98	160.40	216.00	600.00
DKC 05 Logic oper. & exp. CE	22.57	23.89	16.00	245.00	283.06	184.20	264.00	601.00
DKC 06 Other oper. & exp. CE	15.73	20.17	11.00	417.00	186.88	142.35	156.00	601.00
DKC 07 Priority of oper. & exp. CE	14.82	22.05	6.00	240.00	177.37	158.63	132.00	599.00
DKC 08 User input CE	32.78	27.10	26.00	371.00	384.15	235.62	358.00	901.00
DKC 09 User output CE	40.27	42.65	27.00	337.00	552.63	276.55	588.00	901.00
DKC 10 Conditions CE	39.52	43.54	23.00	340.00	526.58	281.09	544.50	901.00
DKC 11 Loops (for) – Basic CE	41.11	37.62	31.00	349.00	481.56	316.68	432.00	1201.00
DKC 12 Loops (for) CE	48.71	52.54	34.00	742.00	633.24	375.71	625.50	1201.00
DKC 13 Loops (while) – Basic CE	51.28	52.48	37.00	436.00	672.05	389.05	685.00	1201.00
DKC 14 Loops (while) CE	48.36	48.68	36.00	531.00	620.78	368.51	597.00	1201.00
DKC 15 Loops – Basic alg. CE	50.82	52.73	37.00	565.00	644.62	371.81	636.00	1201.00
DKC 16 Functions – Basic CE	37.10	46.58	26.00	1019.00	446.62	317.49	388.00	1201.00
DKC 17 Functions CE	42.95	61.56	18.00	1028.00	596.49	398.88	563.00	1201.00
DKC 19 Lists – Basic CE	39.08	40.56	27.00	657.00	462.35	316.19	420.50	1201.00
DKC 21 Lists CE	43.04	55.77	23.00	707.00	574.74	391.75	527.50	1201.00
DKC 23 Files (incl. Func.) CE	35.96	35.09	30.00	544.00	430.95	301.21	399.50	1200.00
DKC 25 Files (incl. Sorting alg.) CE	42.95	53.56	32.00	1182.00	518.22	356.56	489.00	1201.00

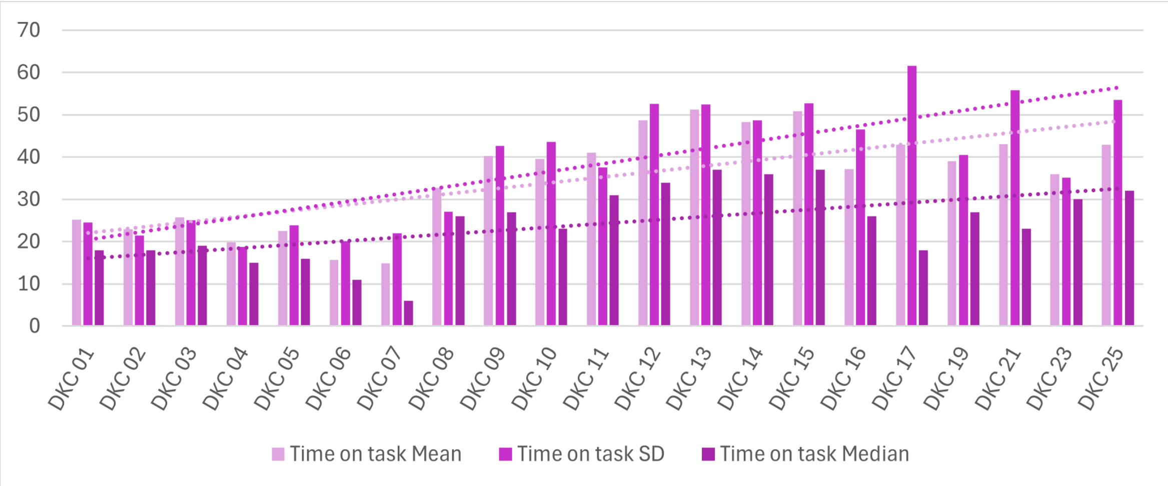

Figure 4.1. The trendline for the time on task feature.

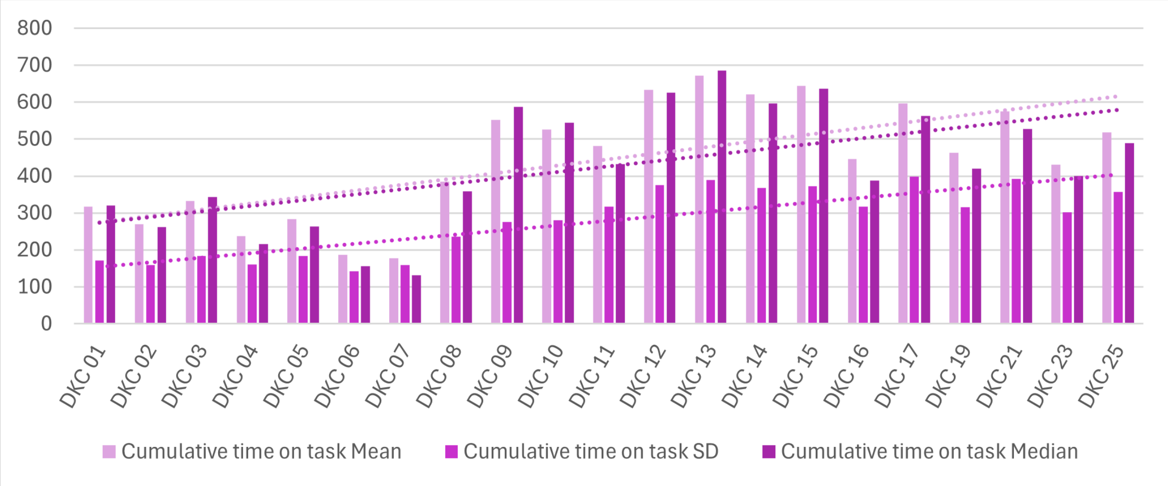

Figure 4.2. The trendline for the cumulative time on task feature.

The total time allocated for each formative assessment (10 minutes for DKC 01-07, 15 minutes for DKC 08-10, and 20 minutes for DKC 11-25) influenced the time on task and cumulative time on task features. Both features showed a noticeable increase as students progressed through the DKCs, indicating that later topics were more cognitively demanding. The growing standard deviations further highlighted an increasing variability in student performance over time, which may be attributed to differences in learning pace, prior knowledge, or levels of engagement. The frequent discrepancy between the median and mean values suggested skewed distributions. Descriptive statistics for the number and cumulative number of code evaluations for each DKC are presented in Table 4.5 and Figures 4.3- 4.4, including the mean, standard deviation, median and maximum values for each feature.

Table 4.5. Descriptive test statistics related to the number and cumulative number of code evaluation features.

Domain Knowledge Component	The number of code evaluations				Cumulative number of code evaluations			
	Mean	SD	Median	Maximum	Mean	SD	Median	Maximum
DKC 01 Basic data types CE	0.15	0.53	0.00	6.00	1.92	3.53	0.00	18.00
DKC 02 Variables CE	0.15	0.54	0.00	6.00	1.76	3.13	0.00	15.00
DKC 03 Arithmetic oper. & exp. CE	0.27	0.75	0.00	7.00	3.78	5.76	0.00	31.00
DKC 04 Relational oper. & exp. CE	0.35	0.73	0.00	8.00	4.25	6.27	1.00	31.00
DKC 05 Logic oper. & exp. CE	0.19	0.67	0.00	7.00	2.65	4.30	0.00	23.00
DKC 06 Other oper. & exp. CE	0.13	0.54	0.00	6.00	1.77	3.38	0.00	18.00
DKC 07 Priority of oper. & exp. CE	0.03	0.23	0.00	3.00	0.49	1.16	0.00	6.00
DKC 08 User input CE	0.32	0.82	0.00	8.00	3.76	5.30	1.00	29.00
DKC 09 User output CE	0.38	0.89	0.00	8.00	5.37	6.64	2.00	32.00
DKC 10 Conditions CE	0.44	0.86	0.00	8.00	5.88	6.74	3.00	30.00
DKC 11 Loops (for) – Basic CE	0.96	1.30	1.00	13.00	10.91	11.46	8.00	61.00
DKC 12 Loops (for) CE	0.64	1.15	0.00	13.00	8.26	9.10	5.00	44.00
DKC 13 Loops (while) – Basic CE	0.89	1.33	0.00	11.00	11.42	11.00	9.00	50.00
DKC 14 Loops (while) CE	0.73	1.13	0.00	13.00	9.13	9.36	7.00	39.00
DKC 15 Loops – Basic alg. CE	0.69	1.03	0.00	10.00	8.23	8.13	6.00	39.00
DKC 16 Functions – Basic CE	0.75	0.96	1.00	13.00	8.31	8.75	6.00	40.00
DKC 17 Functions CE	0.59	1.17	0.00	10.00	8.40	9.12	5.00	41.00
DKC 19 Lists – Basic CE	0.83	0.96	1.00	9.00	9.14	8.81	7.00	46.00
DKC 21 Lists CE	0.66	1.20	0.00	12.00	8.57	9.45	5.00	41.00
DKC 23 Files (incl. Func.) CE	0.70	0.85	1.00	8.00	7.93	8.83	5.00	45.00
DKC 25 Files (incl. Sorting alg.) CE	0.85	1.14	1.00	9.00	9.84	10.43	6.00	48.00

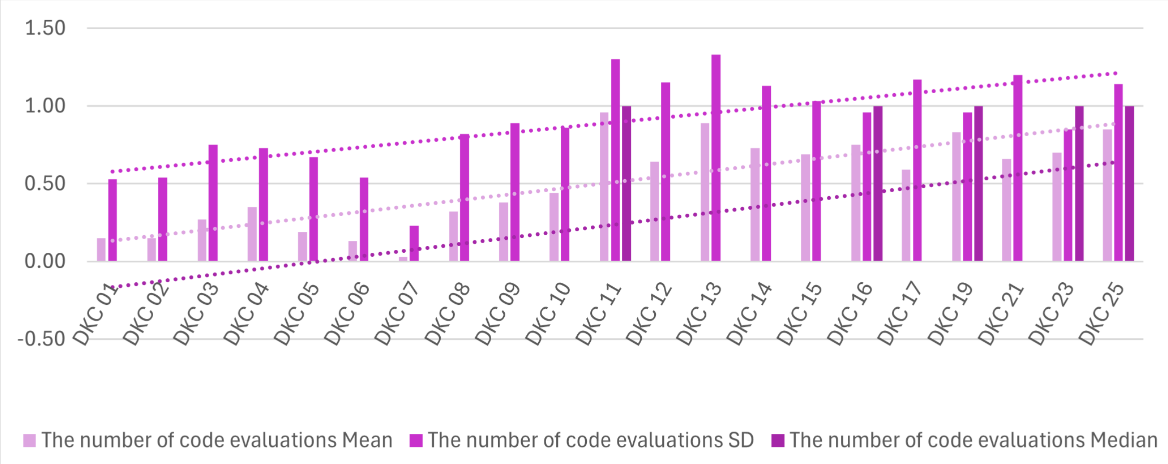

Figure 4.3. The trendline for the number of code evaluation feature.

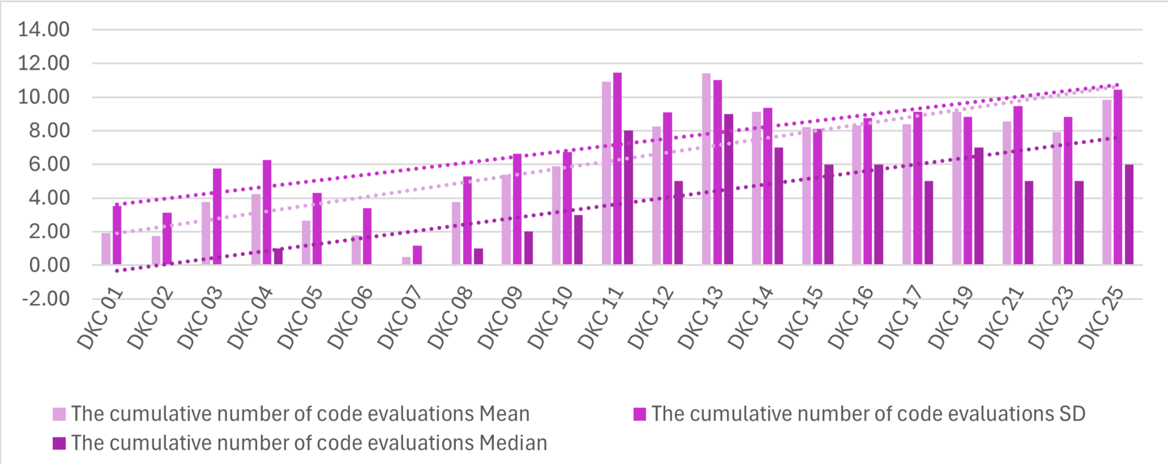

Figure 4.4. The trendline for the cumulative number of code evaluations feature.

The number of code evaluations increased as students advanced through the DKCs, suggesting that these topics were more challenging and required frequent testing and debugging. The growing SDs reflected increasing variability in student behaviour throughout the curriculum. The median cumulative number of code evaluations also rose across DKCs, indicating that students likely evaluated their code more frequently as they progressed. However, the median number of code evaluations for certain DKCs showed that many students did not evaluate their code, particularly during the first week of assessment (DKC 01-07).

Student answers were classified using the Decision tree method, with classification based on extracted features such as time spent on task, cumulative time spent on task, the number of code evaluations and cumulative code evaluations. In addition to arbitrary information gain and entropy criteria, a maximum tree depth was iteratively defined as the minimal depth that effectively differentiated multiple classes across all DKCs. While most DKCs required a depth of 3 to differentiate between multiple classes, a generalized approach of using a depth of 5 proved effective for all DKCs. The example of the resulting decision trees is presented in Figure 4.5.

```
|--- time <= 16.50
|   |--- time <= 1.00
|   |   |--- weights: [103.00, 0.00] class: 0
|   |--- time >  1.00
|   |   |--- cumulative_time <= 600.50
|   |   |   |--- time <= 9.50
|   |   |   |   |--- time <= 7.50
|   |   |   |   |   |--- weights: [137.00, 5.00] class: 0
|   |   |   |   |--- time >  7.50
|   |   |   |   |   |--- weights: [45.00, 0.00] class: 0
|   |   |   |--- time >  9.50
|   |   |   |   |--- cumulative_time <= 10.50
|   |   |   |   |   |--- weights: [0.00, 1.00] class: 1
|   |   |   |   |--- cumulative_time >  10.50
|   |   |   |   |   |--- weights: [76.00, 7.00] class: 0
|   |   |--- cumulative_time >  600.50
|   |   |   |--- cumulative_time <= 835.00
|   |   |   |   |--- time <= 10.50
|   |   |   |   |   |--- weights: [12.00, 2.00] class: 0
|   |   |   |   |--- time >  10.50
|   |   |   |   |   |--- weights: [2.00, 5.00] class: 1
|   |   |   |--- cumulative_time >  835.00
|   |   |   |   |--- time <= 8.50
|   |   |   |   |   |--- weights: [10.00, 2.00] class: 0
|   |   |   |   |--- time >  8.50
|   |   |   |   |   |--- weights: [11.00, 0.00] class: 0
|--- time >  16.50
|   |--- time <= 61.50
|   |   |--- cumulative_time <= 187.00
|   |   |   |--- time <= 24.50
|   |   |   |   |--- time <= 19.50
|   |   |   |   |   |--- weights: [5.00, 0.00] class: 0
|   |   |   |   |--- time >  19.50
|   |   |   |   |   |--- weights: [8.00, 3.00] class: 0
|   |   |   |--- time >  24.50
|   |   |   |   |--- cumulative_time <= 37.50
|   |   |   |   |   |--- weights: [8.00, 3.00] class: 0
|   |   |   |   |--- cumulative_time >  37.50
|   |   |   |   |   |--- weights: [24.00, 43.00] class: 1
|   |   |--- cumulative_time >  187.00
|   |   |   |--- time <= 21.50
|   |   |   |   |--- cumulative_time <= 740.50
|   |   |   |   |   |--- weights: [17.00, 10.00] class: 0
|   |   |   |   |--- cumulative_time >  740.50
|   |   |   |   |   |--- weights: [3.00, 13.00] class: 1
|   |   |   |--- time >  21.50
|   |   |   |   |--- cumulative_time <= 1196.50
|   |   |   |   |   |--- weights: [95.00, 390.00] class: 1
|   |   |   |   |--- cumulative_time >  1196.50
|   |   |   |   |   |--- weights: [4.00, 0.00] class: 0
|   |--- time >  61.50
|   |   |--- time <= 123.50
|   |   |   |--- time <= 75.50
|   |   |   |   |--- time <= 64.50
|   |   |   |   |   |--- weights: [9.00, 4.00] class: 0
|   |   |   |   |--- time >  64.50
|   |   |   |   |   |--- weights: [28.00, 46.00] class: 1
|   |   |   |--- time >  75.50
|   |   |   |   |--- cumulative_time <= 1077.50
|   |   |   |   |   |--- weights: [74.00, 53.00] class: 0
|   |   |   |   |--- cumulative_time >  1077.50
|   |   |   |   |   |--- weights: [1.00, 6.00] class: 1
|   |   |--- time >  123.50
|   |   |   |--- cumulative_time <= 549.00
|   |   |   |   |--- cumulative_time <= 157.50
|   |   |   |   |   |--- weights: [2.00, 3.00] class: 1
|   |   |   |   |--- cumulative_time >  157.50
|   |   |   |   |   |--- weights: [24.00, 2.00] class: 0
|   |   |   |--- cumulative_time >  549.00
|   |   |   |   |--- time <= 133.00
|   |   |   |   |   |--- weights: [4.00, 0.00] class: 0
|   |   |   |   |--- time >  133.00
|   |   |   |   |   |--- weights: [17.00, 13.00] class: 0
```

```
|--- response_evals <= 0.50
|   |--- time <= 17.50
|   |   |--- time <= 1.00
|   |   |   |--- weights: [103.00, 0.00] class: 0
|   |   |--- time >  1.00
|   |   |   |--- time <= 11.50
|   |   |   |   |--- time <= 9.50
|   |   |   |   |   |--- weights: [204.00, 8.00] class: 0
|   |   |   |   |--- time >  9.50
|   |   |   |   |   |--- weights: [32.00, 5.00] class: 0
|   |   |   |--- time >  11.50
|   |   |   |   |--- weights: [58.00, 0.00] class: 0
|   |--- time >  17.50
|   |   |--- cumulative_time <= 201.00
|   |   |   |--- cumulative_evals <= 2.50
|   |   |   |   |--- time <= 82.00
|   |   |   |   |   |--- weights: [42.00, 1.00] class: 0
|   |   |   |   |--- time >  82.00
|   |   |   |   |   |--- weights: [4.00, 2.00] class: 0
|   |   |   |--- cumulative_evals >  2.50
|   |   |   |   |--- weights: [0.00, 1.00] class: 1
|   |   |--- cumulative_time >  201.00
|   |   |   |--- time <= 104.00
|   |   |   |   |--- cumulative_time <= 215.50
|   |   |   |   |   |--- weights: [0.00, 4.00] class: 1
|   |   |   |   |--- cumulative_time >  215.50
|   |   |   |   |   |--- weights: [108.00, 51.00] class: 0
|   |   |   |--- time >  104.00
|   |   |   |   |--- weights: [17.00, 0.00] class: 0
|--- response_evals >  0.50
|   |--- time <= 61.50
|   |   |--- cumulative_time <= 1199.00
|   |   |   |--- time <= 42.50
|   |   |   |   |--- cumulative_evals <= 14.50
|   |   |   |   |   |--- weights: [16.00, 121.00] class: 1
|   |   |   |   |--- cumulative_evals >  14.50
|   |   |   |   |   |--- weights: [5.00, 179.00] class: 1
|   |   |   |--- time >  42.50
|   |   |   |   |--- cumulative_evals <= 40.50
|   |   |   |   |   |--- weights: [23.00, 130.00] class: 1
|   |   |   |   |--- cumulative_evals >  40.50
|   |   |   |   |   |--- weights: [1.00, 0.00] class: 0
|   |   |--- cumulative_time >  1199.00
|   |   |   |--- weights: [2.00, 0.00] class: 0
|   |--- time >  61.50
|   |   |--- time <= 77.50
|   |   |   |--- response_evals <= 2.50
|   |   |   |   |--- cumulative_time <= 770.50
|   |   |   |   |   |--- weights: [11.00, 29.00] class: 1
|   |   |   |   |--- cumulative_time >  770.50
|   |   |   |   |   |--- weights: [11.00, 6.00] class: 0
|   |   |   |--- response_evals >  2.50
|   |   |   |   |--- cumulative_time <= 565.00
|   |   |   |   |   |--- weights: [1.00, 5.00] class: 1
|   |   |   |   |--- cumulative_time >  565.00
|   |   |   |   |   |--- weights: [0.00, 8.00] class: 1
|   |   |--- time >  77.50
|   |   |   |--- cumulative evals <= 13.50
|   |   |   |   |--- response_evals <= 4.50
|   |   |   |   |   |--- weights: [46.00, 29.00] class: 0
|   |   |   |   |--- response_evals >  4.50
|   |   |   |   |   |--- weights: [7.00, 0.00] class: 0
|   |   |   |--- cumulative_evals >  13.50
|   |   |   |   |--- time <= 249.00
|   |   |   |   |   |--- weights: [28.00, 29.00] class: 1
|   |   |   |   |--- time >  249.00
|   |   |   |   |   |--- weights: [0.00, 3.00] class: 1
```

Figure 4.5. The example of Decision trees for DKC 25 CE.

The left Decision tree on Figure 4.5 primarily splits on time on task as its root feature, with a key threshold at 16.50. If $time \leq 1.00$, the classifications belong to Class 0 (103 versus 0), making this a strong model boundary. On the other hand, further divisions are made based on $cumulative\_time$, and one of the most significant differences occurs at $21.5 < time \leq 61.50$ and $187.00 < cumulative\_time \leq 1196.50$, where Class 1 dominates (95 versus 390). This classification suggests that shorter time on task and lower cumulative time tend to classify instances as Class 0, whereas longer time on task and higher cumulative time shift towards Class 1.

The right Decision tree on Figure 4.5 is structured around the number of code evaluations (response_evals feature), with an initial split at 0.50, suggesting its binary use. If $time \leq 1.00$, the majority of cases belong to Class 0 (103 versus 0). A major differentiation in the number of cases also occurs at $time \leq 9.50$, where weights strongly support Class 0 (204 versus 8). On the other hand, if $response\_evals > 0.50$, $time \leq 42.50$ and $cumulative\_time \leq 1199.00$, weights significantly bias the classification towards Class 1 (21 versus 300). For longer response $time > 42.50$, if $cumulative\_evals \leq 40.50$, weights make a strong case for Class 1 (5 versus 179). This classification indicates that a lower number of code evaluations and a shorter time on task lead to Class 0, while more evaluations and a longer time on task shift towards Class 1 classification.

The resulting classification models were cross-validated using training subsets (70%) and evaluated on testing subsets (30%). The Decision tree approach analysed the classification of each answer opportunity (question) in the assessment based on the extracted features.

Standard performance metrics—AUC, F1, Precision, Recall, and Accuracy—assessed classification model outcomes. The AUC metric evaluated the model's ability to distinguish between classes, while the F1 Score balanced Precision and Recall, providing an overall measure of accuracy. Precision reflected the proportion of true positive predictions among all positive predictions, while Recall represented the proportion of true positives identified among all actual positives. Accuracy measured the proportion of correct predictions among all predictions. All metrics ranged from 0 to 1, with higher values indicating better performance.

Two classification approaches were used based on the features extracted. The first classification model considered time on task features (time on a task and the cumulative time on a task), while the second classification additionally considered code evaluation features (the number of code evaluations and cumulative code evaluations). The results for each DKC are presented in Tables 4.6 and 4.7. A heatmap was used for each metric to facilitate visual comparison. Italicized values indicate improvements in cross-validation results, while bolded values in Table 4.7 reflect improvements in time-based classification results compared to those in Table 4.6.

Table 4.6. The classifications based on time on task features.

DKC	Cross-validation					Evaluation				
	AUC	F1	Precision	Recall	Accuracy	AUC	F1	Precision	Recall	Accuracy
DKC 01 CE	0.69379	0.59007	0.53853	0.66981	0.63780	0.64679	0.53061	0.70507	0.84530	0.69024
DKC 02 CE	0.70827	0.73509	0.62396	0.89637	0.65796	0.62886	0.69531	0.75460	0.38801	0.62500
DKC 03 CE	0.74470	0.59525	0.52872	0.68987	0.69451	0.68323	0.58252	0.83032	0.64789	0.67050
DKC 04 CE	0.74229	0.65786	0.64245	0.67937	0.66990	0.66353	0.63946	0.74312	0.61132	0.65584
DKC 05 CE	0.74439	0.01000	0.10000	0.00526	0.81410	0.50948	0.06122	0.79767	0.98563	0.78995
DKC 06 CE	0.78443	0.55050	0.53417	0.57500	0.72754	0.70240	0.59732	0.80984	0.79935	0.73684
DKC 07 CE	0.83555	0.09582	0.13857	0.08889	0.89133	0.49874	0.00000	0.92941	0.99747	0.92723
DKC 08 CE	0.72407	0.76304	0.72602	0.80629	0.70123	0.67910	0.76673	0.59396	0.59000	0.70316
DKC 09 CE	0.79998	0.43270	0.49127	0.39774	0.77357	0.64403	0.45045	0.83495	0.86432	0.76357
DKC 10 CE	0.75554	0.60703	0.54520	0.70991	0.68794	0.58243	0.38515	0.72838	0.84806	0.67995
DKC 11 CE	0.73642	0.79822	0.74378	0.86339	0.72611	0.69284	0.79921	0.73430	0.50498	0.73228
DKC 12 CE	0.80849	0.57990	0.58516	0.57799	0.74577	0.71291	0.57883	0.85404	0.78159	0.74409
DKC 13 CE	0.76698	0.59046	0.59133	0.59402	0.70238	0.71161	0.62574	0.79701	0.79872	0.73750
DKC 14 CE	0.69790	0.27381	0.34778	0.23274	0.66570	0.49906	0.00000	0.72777	0.99812	0.72678
DKC 15 CE	0.79582	0.72113	0.64736	0.81925	0.73043	0.73888	0.72781	0.80247	0.68421	0.73333
DKC 16 CE	0.76722	0.79098	0.72159	0.87626	0.74794	0.72413	0.77636	0.78261	0.58929	0.73416
DKC 17 CE	0.83468	0.46284	0.47101	0.46824	0.78335	0.69058	0.49219	0.88982	0.89280	0.82094
DKC 19 CE	0.80328	0.85957	0.79197	0.94038	0.80376	0.75033	0.87135	0.85714	0.54795	0.81464
DKC 21 CE	0.86289	0.54121	0.61311	0.49372	0.79652	0.69201	0.55414	0.81952	0.91376	0.79167
DKC 23 CE	0.85986	0.83897	0.78300	0.90720	0.80789	0.78426	0.83794	0.83983	0.66438	0.80088
DKC 25 CE	0.81712	0.74806	0.72441	0.77737	0.76015	0.76210	0.77027	0.79537	0.71280	0.76140

Table 4.7. The classifications based on time on task and code evaluation features.

DKC	Cross-validation					Evaluation				
	AUC	F1	Precision	Recall	Accuracy	AUC	F1	Precision	Recall	Accuracy
DKC 01 CE	0.73437	0.49512	0.63596	0.41938	0.67244	0.66427	0.60163	0.74850	0.69061	0.67003
DKC 02 CE	0.71332	0.70676	0.65475	0.77183	0.66207	0.62142	0.67308	0.69458	0.44479	0.61859
DKC 03 CE	0.76499	0.38449	0.69678	0.34064	0.70938	0.59634	0.33654	0.72557	0.98310	0.73563
DKC 04 CE	0.81552	0.69867	0.70701	0.70533	0.72280	0.73693	0.71715	0.82857	0.65660	0.72511
DKC 05 CE	0.77239	0.17028	0.30598	0.13187	0.80329	0.62021	0.39063	0.83750	0.96264	0.82192
DKC 06 CE	0.79019	0.50387	0.64498	0.43458	0.75474	0.69926	0.58091	0.78729	0.92233	0.77851
DKC 07 CE	0.83155	0.14941	0.28524	0.12222	0.89435	0.49495	0.00000	0.92891	0.98990	0.92019
DKC 08 CE	0.77322	0.76974	0.72690	0.82027	0.70698	0.69889	0.77927	0.61589	0.62000	0.72019
DKC 09 CE	0.81962	0.57170	0.65812	0.51329	0.82838	0.71361	0.56456	0.86570	0.89615	0.81266
DKC 10 CE	0.84690	0.67487	0.85432	0.56144	0.81474	0.75594	0.67273	0.82462	0.94700	0.82609
DKC 11 CE	0.84497	0.81525	0.80069	0.83304	0.76263	0.72800	0.80821	0.72984	0.60133	0.75459
DKC 12 CE	0.90120	0.70167	0.77035	0.64665	0.83407	0.79322	0.70707	0.88153	0.91336	0.84777
DKC 13 CE	0.87544	0.74071	0.79842	0.69429	0.82440	0.79701	0.73814	0.84836	0.88651	0.82361
DKC 14 CE	0.87636	0.74052	0.73049	0.75405	0.83607	0.79273	0.68868	0.89546	0.85178	0.81967
DKC 15 CE	0.89065	0.75339	0.77291	0.74471	0.79193	0.79164	0.75627	0.77602	0.90263	0.80290
DKC 16 CE	0.92283	0.87049	0.91087	0.83492	0.86540	0.88269	0.88385	0.83924	0.91667	0.88017
DKC 17 CE	0.91140	0.64816	0.66888	0.64218	0.86126	0.82383	0.68592	0.94118	0.91122	0.88017
DKC 19 CE	0.91166	0.88318	0.86622	0.90274	0.84782	0.80974	0.87646	0.77295	0.73059	0.83489
DKC 21 CE	0.93135	0.72183	0.76335	0.69757	0.86794	0.83455	0.78209	0.89655	0.96099	0.89137
DKC 23 CE	0.94277	0.88165	0.86107	0.90822	0.86536	0.84711	0.87245	0.84286	0.80822	0.85251
DKC 25 CE	0.89730	0.79778	0.82297	0.78062	0.81880	0.83125	0.82545	0.82060	0.85467	0.83158

The average evaluation metrics for classifications based solely on time on task features were 0.66654 for AUC, 0.55917 for the F1 score, 0.79177 for Precision, 0.74600 for Recall, and 0.74476 for Accuracy. In contrast, when incorporating code evaluation features, the average results improved to 0.73969 for AUC, 0.66386 for the F1 score, 0.81437 for Precision, 0.83100 for Recall, and 0.80229 for Accuracy.

The later DKCs in the course showed more consistent student performance, indicating that the classification models were more effective at capturing the underlying patterns in both cross-validation and evaluation phases. For these DKCs, including code evaluation features led to most metrics improvements. This suggests that the significance of code evaluations may vary depending on the complexity of the DKC and the underlying data.

However, a decline in performance from cross-validation to evaluation results indicates neither approach generalised effectively. This highlights a potential issue of overfitting to the training data. Additionally, the requirement for extensive student engagement across eight weekly tests may have influenced performance outcomes.

Both classification models were further used for the BKT-based cognitive modelling.

4.3 BKT model parameters

Eighteen BKT models were analyzed, including the vanilla model (#01), its forgetting-enhanced version (#02), eight models based on time on task features (#03-#10) and eight models that incorporate both time on task and code evaluations features (#11-#18). The BKT models are described using the probabilities of prior knowledge parameter (Prior), guessing parameter (Guess), slipping parameter (Slip), transitioning to the learned knowledge state parameter (Learn) and forgetting parameter (Forgets). The pyBKT models based on multiple classes include the multiple guess and slip model (multigs), the multiple learn model (multilearn) and the multiple prior knowledge model (multiprior).

Regarding the number of parameters fitted for each BKT model, the vanilla model required four parameters: Prior, Guess, Slip, and Learn. The forgetting-enhanced version of the vanilla model and the multilearn and multiprior models required five parameters. The multigs models considered a total of six parameters. Most models utilized seven parameters (multigs+forgets, multilearn+forgets, multigs+multilearn). The multiprior+forgets models accounted for eight parameters, however, experimental results indicated that only the Default class of the Forget parameter as Class1 and Class2 of the Forget parameter consistently equalled zero. The maximum number of parameters considered was nine, as seen in the multigs+multilearn+forgets models.

The BKT models were fitted using the training subsets and 5 initializations of the EM algorithm. The models were seeded to preserve the replicability and the Guess and Slip parameter probabilities were bounded to 0.5.

The resulting parameter probabilities for each DKC are presented in Tables A1- A21 (Appendix A).

To analyse the resulting BKT parameters, the dispersion metrics were calculated for each model across DKCs 01-25. The results include Minimum (Min), Maximum (Max), Mean, Standard Deviation (SD) and the 75% Median (Med 75%) for each parameter, as detailed in Tables 4.8- 4.12.

Table 4.8. BKT model parameter probabilities – Dispersion metrics for DKCs 01-25 (Prior).

BKT model	BKT parameter				
	Prior				
	Default				
	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.12939	0.65433	0.42214	0.14711	0.53430
#02 vanilla+forgets	0.14437	0.76249	0.49795	0.12661	0.53672
#03 multigs T	0.13579	0.84200	0.55522	0.19523	0.66825
#04 multigs+forgets T	0.15644	0.96975	0.62492	0.20268	0.77104
#05 multilearn T	0.12925	0.65197	0.41928	0.14429	0.54326
#06 multilearn+forgets T	0.13624	0.74252	0.49829	0.13043	0.55069
#07 multigs+multilearn T	0.13575	0.84225	0.54882	0.19607	0.66910
#08 multigs+multilearn+forgets T	0.15541	0.99931	0.66054	0.19277	0.76980
#09 multiprior T					
#10 multiprior+forgets T					
#11 multigs TE	0.13438	0.84072	0.55192	0.18466	0.66916
#12 multigs+forgets TE	0.16066	0.88373	0.64066	0.15766	0.74222
#13 multilearn TE	0.12937	0.65188	0.41984	0.14448	0.53583
#14 multilearn+forgets TE	0.13681	0.76894	0.50421	0.13127	0.54639
#15 multigs+multilearn TE	0.13786	0.84297	0.55412	0.18488	0.66952
#16 multigs+multilearn+forgets TE	0.15886	0.95733	0.65077	0.18161	0.75921
#17 multiprior TE					
#18 multiprior+forgets TE					

Depending on the BKT model, the Guess parameter reflects either a single class probability (Default) or a multi-class probability (Class0 and Class1). The vanilla model yielded a Default probability of 0.10643 (SD 0.07253), which increased to 0.14355 (SD 0.08840, #09 and #17) for multiprior models. This baseline result represented a moderate value compared to the multi-class probabilities. In the multi- class BKT models, Class0 values ranged from 0.01021 (SD 0.01350, #08) to 0.03162 (SD 0.02165, #11), while Class 1 values ranged from 0.08545 (SD 0.09698, #12) to 0.20002 (SD 0.11332, #11).

Similarly, the Slip parameter resulted in either single- or multi-class probabilities. The baseline vanilla model produced a moderate value of 0.30264 (SD 0.13332), with other single-class models reaching the lowest value of 0.19513 (SD 0.10895, #10 and #18). In the multi-class BKT models, Class0 values ranged from 0.45721 (SD 07395, #04) to 0.49914 (SD 0.00329, #15), while Class1 values from 0.15725 (SD 0.07672, #16) to 0.23151 (SD 0.11313, #07).

The baseline vanilla model resulted in a Learn parameter value of 0.00699 (SD 0.00901). Other models also had low values of the Default class, with the highest being 0.04381 (SD 0.02206, #10 and #18). In addition, the multiprior models, which specifically considered student answers at the first opportunity (Class 1 and Class 2), produced higher Learn parameter values, ranging from 0.09026 (SD 0.06764, Class1 of the #17 model) to 0.63626 (SD 0.21095, Class2 of the #18 model).

The Forgets parameter in the baseline vanilla model yielded a value of 0.04263 (SD 0.02697), representing a moderate value compared to both single and multi-class models. In the multi-class models, this parameter ranged from 0.03906 (SD 0.02125, Class1 of the #14 model) to 0.07714 (SD 0.07693, Class0 of the #16).

To further discuss the probabilities of the BKT parameters, dispersion metrics were pre-

BKT model	BKT parameter														
	Guess														
	Default					Guess_Class0					Guess_Class1				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.01019	0.29955	0.10643	0.07253	0.12490										
#02 vanilla+forgets	0.00335	0.26031	0.07771	0.06508	0.09411										
#03 multigs T						0.00000	0.08302	0.02965	0.02231	0.03802	0.00074	0.50000	0.15575	0.15489	0.18963
#04 multigs+forgets T						0.00000	0.05163	0.01084	0.01441	0.01251	0.00000	0.50000	0.10653	0.17166	0.09522
#05 multilearn T	0.01111	0.30726	0.11112	0.07492	0.12453										
#06 multilearn+forgets T	0.00790	0.25798	0.08398	0.06783	0.10303										
#07 multigs+multilearn T						0.00334	0.09423	0.03033	0.02367	0.03881	0.02429	0.50000	0.15754	0.15366	0.19107
#08 multigs+multilearn+forgets T						0.00000	0.05243	0.01021	0.01350	0.01674	0.00000	0.50000	0.09995	0.17217	0.09129
#09 multiprior T	0.01242	0.33327	0.14355	0.08840	0.20401										
#10 multiprior+forgets T	0.01015	0.29982	0.09808	0.07687	0.12120										
#11 multigs TE						0.00906	0.07806	0.03162	0.02165	0.04430	0.01983	0.43987	0.20002	0.11332	0.28661
#12 multigs+forgets TE						0.00000	0.09459	0.01544	0.02274	0.02304	0.00002	0.31008	0.08545	0.09698	0.14291
#13 multilearn TE	0.01066	0.30855	0.11015	0.07441	0.12458										
#14 multilearn+forgets TE	0.00393	0.27051	0.08231	0.06886	0.10015										
#15 multigs+multilearn TE						0.00800	0.07813	0.03101	0.02211	0.04122	0.01425	0.42574	0.19619	0.11762	0.29028
#16 multigs+multilearn+forgets TE						0.00000	0.03273	0.01105	0.01111	0.01833	0.00001	0.33951	0.08660	0.08931	0.15156
#17 multiprior TE	0.01242	0.33327	0.14355	0.08840	0.20401										
#18 multiprior+forgets TE	0.01015	0.29982	0.09808	0.07687	0.12120										

Table 4.9. BKT model parameter probabilities – Dispersion metrics for DKCs 01-25 (Guess).

BKT model	BKT parameters														
	Slip														
	Default					Slip_Class0					Slip_Class1				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.11001	0.50000	0.30264	0.13332	0.39336										
#02 vanilla+forgets	0.08833	0.50000	0.24739	0.11902	0.34000										
#03 multigs T						0.37107	0.50000	0.49309	0.02818	0.50000	0.00000	0.44436	0.23097	0.11274	0.28440
#04 multigs+forgets T						0.26689	0.50000	0.45721	0.07395	0.50000	0.00000	0.43407	0.19710	0.11400	0.26487
#05 multilearn T	0.10964	0.50000	0.29989	0.13568	0.39452										
#06 multilearn+forgets T	0.08033	0.50000	0.24329	0.12100	0.33374										
#07 multigs+multilearn T						0.36724	0.50000	0.49268	0.02896	0.50000	0.00000	0.44426	0.23151	0.11313	0.28069
#08 multigs+multilearn+forgets T						0.27090	0.50000	0.47939	0.05525	0.50000	0.00000	0.40452	0.20343	0.09764	0.26481
#09 multiprior T	0.07679	0.50000	0.25432	0.12000	0.35091										
#10 multiprior+forgets T	0.03298	0.48790	0.19513	0.10895	0.25075										
#11 multigs TE						0.47322	0.50000	0.49842	0.00594	0.50000	0.07287	0.34616	0.19231	0.08697	0.26791
#12 multigs+forgets TE						0.25700	0.50000	0.48257	0.05446	0.50000	0.05310	0.32816	0.16979	0.08367	0.25695
#13 multilearn TE	0.10972	0.50000	0.30037	0.13522	0.39143										
#14 multilearn+forgets TE	0.08443	0.50000	0.24215	0.12033	0.33338										
#15 multigs+multilearn TE						0.48510	0.50000	0.49914	0.00329	0.50000	0.07287	0.33943	0.19319	0.08595	0.26813
#16 multigs+multilearn+forgets TE						0.42920	0.50000	0.49070	0.01799	0.50000	0.04565	0.31405	0.15725	0.07672	0.21576
#17 multiprior TE	0.07679	0.50000	0.25432	0.12000	0.35091										
#18 multiprior+forgets TE	0.03298	0.48790	0.19513	0.10895	0.25075										

Table 4.10. BKT model parameter probabilities – Dispersion metrics for DKCs 01-25 (Slip).

BKT model	BKT parameters														
	Learn														
	Default					Learn_Class0 (Class1 multiprior)					Learn_Class1 (Class2 multiprior)				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.00009	0.03333	0.00699	0.00859	0.00905										
#02 vanilla+forgets	0.00058	0.05054	0.01824	0.01289	0.02348										
#03 multigs T	0.00004	0.03032	0.00759	0.00883	0.01033										
#04 multigs+forgets T	0.00026	0.05926	0.02707	0.01879	0.03861										
#05 multilearn T						0.00000	0.02238	0.00498	0.00670	0.00793	0.00000	0.04011	0.00895	0.01097	0.01604
#06 multilearn+forgets T						0.00031	0.04093	0.01363	0.01164	0.01593	0.00004	0.06898	0.02159	0.01777	0.02919
#07 multigs+multilearn T						0.00011	0.03636	0.00716	0.00948	0.00997	0.00004	0.04116	0.01192	0.01372	0.02044
#08 multigs+multilearn+forgets T						0.00002	0.07741	0.02067	0.01972	0.02575	0.00005	0.14602	0.03689	0.03943	0.04794
#09 multiprior T	0.00151	0.06641	0.02229	0.01515	0.02924	0.00769	0.30279	0.14828	0.08107	0.21291	0.13082	1.00000	0.48460	0.31470	0.53676
#10 multiprior+forgets T	0.00562	0.08814	0.04381	0.02206	0.05330	0.05752	0.37990	0.20710	0.08644	0.25404	0.29755	1.00000	0.56551	0.26655	0.70026
#11 multigs TE	0.00012	0.02705	0.00759	0.00772	0.01098										
#12 multigs+forgets TE	0.00027	0.15113	0.02471	0.03313	0.02686										
#13 multilearn TE						0.00000	0.02586	0.00540	0.00703	0.00898	0.00000	0.04130	0.00868	0.01215	0.01137
#14 multilearn+forgets TE						0.00063	0.03460	0.01468	0.01054	0.02341	0.00003	0.07706	0.02648	0.02043	0.03590
#15 multigs+multilearn TE						0.00005	0.03512	0.00752	0.00898	0.00948	0.00001	0.03381	0.01004	0.01116	0.01492
#16 multigs+multilearn+forgets TE						0.00001	0.13927	0.03102	0.03246	0.03075	0.00004	0.24939	0.05189	0.05749	0.05256
#17 multiprior TE	0.00151	0.06641	0.02229	0.01515	0.02924	0.01457	0.30279	0.09026	0.06764	0.13637	0.21485	1.00000	0.52294	0.26051	0.66216
#18 multiprior+forgets TE	0.00562	0.08814	0.04381	0.02206	0.05330	0.05421	0.30581	0.13465	0.06466	0.15046	0.30838	1.00000	0.63626	0.21095	0.77862

Table 4.11. BKT model parameter probabilities – Dispersion metrics for DKCs 01-25 (Learn).

BKT model	BKT parameters														
	Forgets														
	Default					Forgets_Class0 (Class 1 multiprior)					Forgets_Class1 (Class 2 multiprior)				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla															
#02 vanilla+forgets	0.01389	0.09973	0.04263	0.02697	0.06279										
#03 multigs															
#04 multigs+forgets	0.00498	0.14415	0.04710	0.03991	0.06039										
#05 multilearn															
#06 multilearn+forgets						0.00563	0.09902	0.04716	0.02526	0.06314	0.01337	0.36107	0.06190	0.07883	0.06967
#07 multigs+multilearn															
#08 multigs+multilearn+forgets						0.01118	0.13064	0.05154	0.03344	0.06496	0.00562	0.35919	0.05929	0.07940	0.07013
#09 multiprior															
#10 multiprior+forgets	0.02005	0.12528	0.06179	0.03741	0.09869										
#11 multigs															
#12 multigs+forgets	0.00319	0.09740	0.03899	0.03111	0.06502										
#13 multilearn															
#14 multilearn+forgets						0.01077	0.14894	0.05761	0.03533	0.07316	0.01206	0.07399	0.03906	0.02125	0.05651
#15 multigs+multilearn															
#16 multigs+multilearn+forgets						0.01342	0.30250	0.07714	0.07693	0.08601	0.00493	0.19515	0.04582	0.04960	0.06008
#17 multiprior															
#18 multiprior+forgets	0.02005	0.12528	0.06179	0.03741	0.09869										

Table 4.12. BKT model parameter probabilities – Dispersion metrics for DKCs 01-25 (Forgets).

sented for specific model groups based on parameter usage. These groups included the multigs model group (#03, #04, #07, #08, #11, #12, #15, #16), the multilearn group (#05, #06, #07, #08, #13, #14, #15, #16) and the multiprior group (#9, #10, #17, #18).

Additionally, model groups were analysed based on the features used in student answer classification, comprising the time-on-task group (#03 - #10) and the time- and evaluation-based group (#11 - #18).

The results from these model groups were compared to the baseline vanilla BKT model (#01) and its forgetting-enhanced version (#02). The analysis included Minimum (Min), Maximum (Max), Mean, Standard Deviation (SD) and the 75% Median (Med 75%) metrics. The results for each model group across DKCs 01-25 are presented in Tables 4.13- 4.17.

Table 4.13. BKT model parameter probabilities – Dispersion metrics for BKT model groups (Prior).

BKT model	**BKT parameter**				
	Prior				
	Default				
	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.12939	0.65433	0.42214	0.14711	0.53430
#02 vanilla+forgets	0.14437	0.76249	0.49795	0.12661	0.53672
Multigs models	0.13438	0.99931	0.59837	0.18933	0.73695
Multilearn models	0.12925	0.99931	0.53198	0.18319	0.65052
Multiprior models					
Time-based m.	0.12925	0.99931	0.55118	0.19272	0.67093
Time- and evals-based m.	0.12937	0.95733	0.55359	0.18042	0.65903

Compared to the baseline vanilla model, which exhibited the lowest Prior parameter value, the multigs model group showed the highest mean Prior value of 0.59837 (SD 0.18933). The time-based and time- and evaluation-based model groups yielded a similar moderate mean of 0.55118 (SD 0.19272) and 0.55359 (SD 0.18042). The Prior parameter varied significantly, especially across the two feature-based model groups.

While the baseline vanilla model produced a moderate Guess parameter value, the multiprior model group showed slightly higher values. The multigs and multilearn model groups distinguished Class0 with a lower Guess probability (0.02127 and 0.02065, SD 0.02135 and SD 0.02069) and Class1 with a significantly higher Guess probability (0.13600, SD 0.14155 and SD 0.14107). The feature-based model groups followed a similar pattern of lower Class0 and higher Class1 probabilities as seen in the multigs and multilearn group models.

For the Slip parameter, the multi-class model groups and the feature-based model group defined Class0 as significantly higher and Class1 as lower than the baseline vanilla model. The Slip parameter reached a maximum value of 0.5 across all model groups.

The Learn parameter probabilities were generally low across all model groups, often below 0.04, reflecting knowledge acquisition's slow and incremental nature. While the baseline vanilla model showed the lowest value of 0.00699 (SD 0.00859), other models produced Default probabilities ranging from 0.01674 (multigs model group, SD 0.02164) to 0.03305 (multiprior model group, SD 0.02150). In the multi-class models, Class0 ranged from 0.01313 (multilearn model group, SD 0.01750) to 0.14507 (multiprior model group,

BKT model	BKT parameter														
	Guess														
	Default					Guess_Class0					Guess_Class1				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.01019	0.29955	0.10643	0.07253	0.12490										
#02 vanilla+forgets	0.00335	0.26031	0.07771	0.06508	0.09411										
Multigs models						0.00000	0.09459	0.02127	0.02135	0.03198	0.00000	0.50000	0.13600	0.14155	0.19795
Multilearn models	0.00393	0.30855	0.09689	0.07162	0.11972	0.00000	0.09423	0.02065	0.02069	0.03038	0.00000	0.50000	0.13600	0.14107	0.19772
Multiprior models	0.01015	0.33327	0.12082	0.08448	0.17335										
Time-based m.	0.00790	0.33327	0.10918	0.07912	0.13889	0.00000	0.09423	0.02026	0.02109	0.03275	0.00000	0.50000	0.12994	0.16261	0.14348
Time- and evals-based m.	0.00393	0.33327	0.10852	0.07935	0.13889	0.00000	0.09459	0.02228	0.02168	0.03096	0.00001	0.43987	0.14206	0.11746	0.22447

Table 4.14. BKT model parameter probabilities – Dispersion metrics for BKT model groups (Guess).

BKT model	BKT parameters														
	Slip														
	Default					Slip_Class0					Slip_Class1				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.11001	0.50000	0.30264	0.13332	0.39336										
#02 vanilla+forgets	0.08833	0.50000	0.24739	0.11902	0.34000										
Multigs models						0.25700	0.50000	0.48665	0.04221	0.50000	0.00000	0.44436	0.19694	0.09840	0.26797
Multilearn models	0.08033	0.50000	0.27143	0.12920	0.35410	0.27090	0.50000	0.49048	0.03270	0.50000	0.00000	0.44426	0.19634	0.09641	0.26564
Multiprior models	0.03298	0.50000	0.22472	0.11639	0.30787										
Time-based m.	0.03298	0.50000	0.24816	0.12529	0.33485	0.26689	0.50000	0.48059	0.05159	0.50000	0.00000	0.44436	0.21575	0.10873	0.27552
Time- and evals-based m.	0.03298	0.50000	0.24799	0.12508	0.33458	0.25700	0.50000	0.49270	0.02915	0.50000	0.04565	0.34616	0.17813	0.08333	0.25705

Table 4.15. BKT model parameter probabilities – Dispersion metrics for BKT model groups (Slip).

BKT model	BKT parameters														
	Learn														
	Default					Learn_Class0 (Class1 multiprior)					Learn_Class1 (Class2 multiprior)				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla	0.00009	0.03333	0.00699	0.00859	0.00905										
#02 vanilla+forgets	0.00058	0.05054	0.01824	0.01289	0.02348										
Multigs models	0.00004	0.15113	0.01674	0.02164	0.02193	0.00001	0.13927	0.01659	0.02211	0.02159	0.00001	0.24939	0.02768	0.03947	0.03573
Multilearn models						0.00000	0.13927	0.01313	0.01750	0.01821	0.00000	0.24939	0.02206	0.03093	0.02873
Multiprior models	0.00151	0.08814	0.03305	0.02150	0.04269	0.00769	0.37990	0.14507	0.08519	0.20381	0.13082	1.00000	0.55233	0.26696	0.70118
Time-based m.	0.00004	0.08814	0.02519	0.02111	0.03573	0.00000	0.37990	0.06697	0.09406	0.10202	0.00000	1.00000	0.18824	0.29218	0.31253
Time- and evals-based m.	0.00012	0.15113	0.02460	0.02489	0.03397	0.00000	0.30581	0.04726	0.06312	0.06646	0.00000	1.00000	0.20938	0.29829	0.39088

Table 4.16. BKT model parameter probabilities – Dispersion metrics for BKT model groups (Learn).

BKT model	BKT parameters														
	Forgets														
	Default					Forgets_Class0 (Class 1 multiprior)					Forgets_Class1 (Class 2 multiprior)				
	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%	Min	Max	Mean	SD	Med 75%
#01 vanilla															
#02 vanilla+forgets	0.01389	0.09973	0.04263	0.02632	0.06279										
Multigs models	0.00319	0.14415	0.04304	0.03558	0.06329	0.01118	0.30250	0.06434	0.06000	0.07735	0.00493	0.35919	0.05255	0.06574	0.06481
Multilearn models						0.00563	0.30250	0.05836	0.04778	0.06936	0.00493	0.36107	0.05152	0.06171	0.06405
Multiprior models	0.02005	0.12528	0.06179	0.03695	0.09869										
Time-based m.	0.00498	0.14415	0.05444	0.03892	0.08564	0.00000	0.13064	0.03290	0.03346	0.05734	0.00000	0.36107	0.04040	0.06977	0.05422
Time- and evals-based m.	0.00319	0.12528	0.05039	0.03589	0.07144	0.00000	0.30250	0.04492	0.05832	0.06680	0.00000	0.19515	0.02829	0.03679	0.04735

Table 4.17. BKT model parameter probabilities – Dispersion metrics for BKT model groups (Forgets).

SD 0.08519), while Class1 ranged from 0.02206 (multigs model group, SD 0.03093) to 0.55233 (multiprior model group, SD 0.26696).

The Forgets parameter probability averaged up to 0.06434 (SD 0.06000) for Class0 of the multigs model group.

The analysis of eighteen enhanced BKT models revealed that more complex models, incorporating features such as time spent on the task and the number of code evaluations, captured finer details in student learning behaviours. These advanced models effectively differentiated students' patterns of prior knowledge, guessing, slipping, learning, and forgetting, providing clearer insights into the process of knowledge acquisition. Both feature-based model groups (time-based group and time- and evals-based group) effectively distinguished between the educational contexts in which students completed tasks. Students classified in Class0 were generally less likely to guess correctly compared to those in Class1. Similarly, students in Class0 were more prone to making errors despite knowing the material, in contrast to their counterparts in Class1. The Learn parameter for Class0 was consistently lower than that for Class1, indicating a more gradual learning progression for students in Class0.

4.4 Prediction of student performance

Regarding student performance prediction, the BKT models were cross-validated using a training subset consisting of 70% of student-level data and evaluated on a testing subset comprising the remaining 30%. These student-level subsets ensured that each student's data were utilized for training or testing purposes.

The 5-fold cross-validation, which involved combinations of BKT parameter fittings and model evaluations, was applied to the training subset to explore the potential parameter space. The evaluation of models based on the BKT parameters and the unseen data in the testing subset ensured that the models could generalize effectively.

The results for each DKC were reported using RMSE, AUC, and Accuracy metrics and can be found in Tables B1- B21 (Appendix B).

To discuss the prediction results, the average values across DKCs for each BKT model are presented in Table 4.18. The metrics include RMSE Average (SD), AUC Average (SD) and Accuracy Average (SD), and are visualized using heatmaps.

Furthermore, the average prediction results across DKCs are presented for parameter-based and feature-based model groups, including multigs, multilearn, multiprior, time-based group and time- and evaluation-based group in Table 4.19. Heatmaps were used for each metric, with baseline results repeated for comparison.

The enhanced multigs+forgets (#12) and multigs+multilearn+forgets (#16) showed the best average performance across DKCs. The evaluation results for model #12 included an RMSE of 0.35456 (SD 0.04347), AUC of 0.88248 (SD 0.04830), and Accuracy of 0.82999 (SD 0.05266). Similarly, model #16 achieved an RMSE of 0.35160 (SD 0.04329), AUC of

BKT model	Cross-validation						Evaluation					
	RMSE Average	SD	AUC Average	SD	Accuracy Average	SD	RMSE Average	SD	AUC Average	SD	Accuracy Average	SD
#01 vanilla	0.38476	0.05072	0.80740	0.06059	0.77942	0.05782	0.39401	0.03766	0.81570	0.06777	0.77260	0.04908
#02 vanilla+forgets	0.37216	0.05080	0.83268	0.05424	0.79606	0.05292	0.37857	0.04079	0.84404	0.05365	0.78993	0.05310
#03 multigs T	0.37630	0.05474	0.83065	0.05875	0.80251	0.05839	0.38871	0.03798	0.83427	0.05678	0.79694	0.04408
#04 multigs+forgets T	0.36594	0.05423	0.85260	0.05602	0.80414	0.05975	0.37311	0.04034	0.86193	0.04848	0.80009	0.04527
#05 multilearn T	0.38473	0.04996	0.80744	0.06073	0.78039	0.05733	0.39363	0.03773	0.81622	0.06812	0.77369	0.04865
#06 multilearn+forgets T	0.37283	0.05076	0.83170	0.05388	0.79478	0.05390	0.37857	0.04043	0.84264	0.05272	0.78955	0.05373
#07 multigs+multilearn T	0.37614	0.05450	0.83203	0.05890	0.80384	0.05773	0.38824	0.03782	0.83576	0.05670	0.79752	0.04310
#08 multigs+multilearn+forgets T	0.36257	0.05370	0.85214	0.05420	0.80679	0.05827	0.37114	0.03940	0.85979	0.04673	0.80125	0.04500
#09 multiprior T	0.39352	0.05410	0.78797	0.06213	0.77231	0.05918	0.40264	0.04100	0.80249	0.05956	0.76141	0.05349
#10 multiprior+forgets T	0.37960	0.05332	0.80787	0.06003	0.78662	0.05535	0.38636	0.04342	0.82227	0.05521	0.78045	0.05494
#11 multigs TE	0.36135	0.05069	0.85245	0.05757	0.82880	0.05698	0.36869	0.04007	0.86327	0.05699	0.83101	0.04353
#12 multigs+forgets TE	0.35033	0.05172	0.86909	0.05580	0.83146	0.06008	0.35456	0.04347	0.88248	0.04830	0.82999	0.05266
#13 multilearn TE	0.38473	0.04999	0.80753	0.06063	0.78014	0.05765	0.39376	0.03775	0.81594	0.06768	0.77354	0.04869
#14 multilearn+forgets TE	0.37291	0.05075	0.83232	0.05317	0.79481	0.05261	0.37858	0.04063	0.84487	0.05336	0.78953	0.05336
#15 multigs+multilearn TE	0.36142	0.05075	0.85243	0.05742	0.82798	0.05707	0.36872	0.04015	0.86328	0.05669	0.83100	0.04389
#16 multigs+multilearn+forgets TE	0.34775	0.05141	0.86691	0.05648	0.83082	0.05909	0.35160	0.04329	0.88108	0.04899	0.83048	0.05037
#17 multiprior TE	0.39168	0.05416	0.79044	0.06221	0.77621	0.05907	0.40096	0.04140	0.80435	0.06129	0.76443	0.05387
#18 multiprior+forgets TE	0.37725	0.05351	0.80991	0.06006	0.79044	0.05442	0.38420	0.04438	0.82404	0.05647	0.78475	0.05633

Table 4.18. Prediction of student performance – Average results (DKCs 01-25) per BKT model.

BKT model	Cross-validation						Evaluation					
	RMSE Average	SD	AUC Average	SD	Accuracy Average	SD	RMSE Average	SD	AUC Average	SD	Accuracy Average	SD
#01 vanilla	0.38476	0.05072	0.80740	0.06059	0.77942	0.05782	0.39401	0.03766	0.81570	0.06777	0.77260	0.04908
#02 vanilla+forgets	0.37216	0.05080	0.83268	0.05424	0.79606	0.05292	0.37857	0.04079	0.84404	0.05365	0.78993	0.05310
Multigs models	0.36273	0.05272	0.85104	0.05689	0.81704	0.05842	0.37060	0.04031	0.86023	0.05246	0.81478	0.04599
Multilearn models	0.37039	0.05148	0.83531	0.05693	0.80244	0.05671	0.37803	0.03965	0.84495	0.05637	0.79832	0.04835
Multiprior models	0.38551	0.05377	0.79905	0.06111	0.78139	0.05700	0.39354	0.04255	0.81329	0.05813	0.77276	0.05466
Time-based m.	0.37645	0.05316	0.82530	0.05808	0.79392	0.05749	0.38530	0.03977	0.83442	0.05554	0.78761	0.04853
Time- and evals-based m.	0.36967	0.05181	0.83211	0.05815	0.80525	0.05692	0.37638	0.04162	0.84462	0.05611	0.80169	0.05085

Table 4.19. Prediction of student performance – Average results (DKCs 01-25) per BKT model group.

0.88108 (SD 0.04899), and Accuracy of 0.83048 (SD 0.05037).

Among the broader model groups, the multigs group yielded the best average performance, with an RMSE of 0.37060 (SD 0.04031), AUC of 0.86023 (SD 0.05246) and Accuracy of 0.81478 (SD 0.04599). Also, both time- and evaluation-based groups overperformed the baseline vanilla BKT model.

The slightly lower performance observed during evaluation compared to cross-validation indicated potential overfitting in the models, suggesting that further tuning or additional data was necessary to improve generalization. While the trend was consistent across DKCs, some topics appeared to benefit more from including time on task and code evaluation features. This variation may reflect differences in the nature of the content or patterns of student interaction specific to these DKCs.

To quantify the statistical significance of the differences in RMSE and AUC prediction results across DKCs, the nonparametric Wilcoxon Signed Rank Test is employed [122]. This test evaluates whether the two paired groups differ significantly in their medians and assumes the null hypothesis that both groups have identical distributions. The W-value denotes the range in magnitude, and a p-value below 0.05 indicates a statistically significant difference, as presented in bold in Tables 4.20 and 4.21.

Table 4.20. Prediction of student performance (RMSE) – Wilcoxon Signed Rank Test results.

BKT model	**W-value**	**p-value**
#01 vanilla and #02 vanilla+forgets	6	**SS ($p < 0.05$)**
#01 vanilla and #03 multigs T	64	NSS at $p < 0.05$
#01 vanilla and #04 multigs+forgets T	4	**SS ($p < 0.05$)**
#01 vanilla and #05 multilearn T	62	NSS at $p < 0.05$
#01 vanilla and #06 multilearn+forgets T	3	**SS ($p < 0.05$))**
#01 vanilla and #07 multigs+multilearn T	57	NSS at $p < 0.05$
#01 vanilla and #08 multigs+multilearn+forgets T	4	**SS ($p < 0.05$)**
#01 vanilla and #09 multiprior T	19	**SS ($p < 0.05$)**
#01 vanilla and #10 multiprior+forgets T	60	NSS at $p < 0.05$
#01 vanilla and #11 multigs TE	3	**SS ($p < 0.05$)**
#01 vanilla and #12 multigs+forgets TE	1	**SS ($p < 0.05$)**
#01 vanilla and #13 multilearn TE	64	NSS at $p < 0.05$
#01 vanilla and #14 multilearn+forgets TE	5	**SS ($p < 0.05$)**
#01 vanilla and #15 multigs+multilearn TE	3	**SS ($p < 0.05$)**
#01 vanilla and #16 multigs+multilearn+forgets TE	0	**SS ($p < 0.05$)**
#01 vanilla and #17 multiprior TE	26	**SS ($p < 0.05$)**
#01 vanilla and #18 multiprior+forgets TE	43	**SS ($p < 0.05$)**

Table 4.21. Prediction of student performance (AUC) – Wilcoxon Signed Rank Test results.

BKT model	W-value	p-value
#01 vanilla and #02 vanilla+forgets	4	**SS ($p < 0.05$)**
#01 vanilla and #03 multigs T	25	**SS ($p < 0.05$)**
#01 vanilla and #04 multigs+forgets T	1	**SS ($p < 0.05$)**
#01 vanilla and #05 multilearn T	67	NSS at $p < 0.05$
#01 vanilla and #06 multilearn+forgets T	7	**SS ($p < 0.05$)**
#01 vanilla and #07 multigs+multilearn T	16	**SS ($p < 0.05$)**
#01 vanilla and #08 multigs+multilearn+forgets T	1	**SS ($p < 0.05$)**
#01 vanilla and #09 multiprior T	19	**SS ($p < 0.05$)**
#01 vanilla and #10 multiprior+forgets T	96	NSS at $p < 0.05$
#01 vanilla and #11 multigs TE	0	**SS ($p < 0.05$)**
#01 vanilla and #12 multigs+forgets TE	1	**SS ($p < 0.05$)**
#01 vanilla and #13 multilearn TE	70	NSS at $p < 0.05$
#01 vanilla and #14 multilearn+forgets TE	3	**SS ($p < 0.05$)**
#01 vanilla and #15 multigs+multilearn TE	0	**SS ($p < 0.05$)**
#01 vanilla and #16 multigs+multilearn+forgets TE	1	**SS ($p < 0.05$)**
#01 vanilla and #17 multiprior TE	24	**SS ($p < 0.05$)**
#01 vanilla and #18 multiprior+forgets TE	82	NSS at $p < 0.05$

The Wilcoxon Signed Rank Test results indicated a statistically significant difference in RMSE for 12 out of 17 BKT models and AUC for 13 out of 17 models compared to the baseline vanilla BKT model. Specifically:

- For RMSE, four time-based BKT models (#03, #05, #07, #10) and one time- and evaluation-based model (#13) showed no significant difference.
- For AUC, two time-based BKT models (#05 and #10) and two time- and evaluation-based models (#13 and #18) showed no significant difference.

This demonstrated that the enhanced BKT models overperformed the vanilla BKT model in predicting student performance. The reduction in RMSE and improvements in AUC and Accuracy metrics indicated that additional features enhanced the model's ability to capture important aspects of student learning and performance trends across different DKSs. This suggested that the added features were relevant to understanding the nuances of student behaviour and knowledge mastery.

4.5 Estimation of knowledge mastery

In the context of BKT-based adaptive assessment, knowledge mastery is estimated using the final mastery probability on a 0-1 scale. In contrast, in traditional assessment contexts, mastery is typically observed as a percentage-based student performance in both formative and summative assessments (midterm and final exams). The relationship between the final adaptive BKT probability and the student performance in formative and summative assessments was analysed using the Pearson correlation (r) and the related p-value. The model's ability to classify students with positive performance (over 50% overall) was additionally evaluated using the F1 score. The mastery estimation was performed using the testing subset and the complete dataset.

The results for each DKC based on the testing subsets are presented in Tables C1- C21 (Appendix C), while the complete dataset results for each DKC are shown in Tables C22-

C42 (Appendix C).

To summarize, average estimations for the testing subsets across DKCs for each BKT model are presented in Table 4.22, which includes a heatmap for the following metrics: Pearson r Average (SD), p-value Average (SD) and the F1 Average (SD).

The average estimation results across DKCs per parameter-based and feature-based model groups were compared to the baseline vanilla BKT model (#01) and its forgetting enhanced version (#02), with these comparisons displayed in a heatmap (Table 4.23).

Based on the average estimation results, the baseline vanilla model yielded a statistically significant Pearson correlation r of 0.83883 (SD 0.02751) for formative assessments, 0.46968 (SD 0.11314) for the midterm exam and 0.40604 (SD 0.09694) for the final exam. Several models overperformed the baseline vanilla model, including multilearn and multiprior models (#05, #09, #13 and #17) in formative correlations and classification, and multiprior models (#09 and #17) in midterm exam correlations. Models with p-values below 0.05 indicated statistically significant correlations.

The results from the BKT model groups showed that neither parameter- nor feature-based generalizations outperformed the baseline vanilla model.

Most BKT models exhibited strong positive correlations with student performance in formative assessments, often with Pearson r values and F1 scores exceeding 0.7. However, there was a noticeable decline in correlation and classification performance for the summative midterm exam, with further drops for the final exam, reflecting reduced estimation accuracy as the assessments become more temporally distant from the learning sessions.

BKT model	Correlation												Classification					
	Formative				Summative - Midterm				Summative - Final				Formative		Summative - Midterm		Summative - Final	
	Pearson r Average	SD	p-value Average	SD	Pearson r Average	SD	p-value Average	SD	Pearson r Average	SD	p-value Average	SD	F1 Average	SD	F1 Average	SD	F1 Average	SD
#01 vanilla	0.83883	0.02751	0.00000	0.00000	0.46968	0.11314	0.00418	0.00856	0.40604	0.09694	0.02072	0.03231	0.76889	0.17352	0.69666	0.08380	0.64589	0.08309
#02 vanilla+forgets	0.81073	0.06144	0.00000	0.00000	0.43974	0.11718	0.01011	0.03128	0.39704	0.13960	0.04378	0.07438	0.45985	0.44594	0.35823	0.34733	0.33388	0.32423
#03 multigs T	0.72156	0.09483	0.00000	0.00000	0.41672	0.11383	0.01280	0.03577	0.35412	0.13114	0.07552	0.13649	0.68096	0.16413	0.69206	0.06129	0.64350	0.06062
#04 multigs+forgets T	0.68503	0.13403	0.00006	0.00023	0.36854	0.11343	0.03500	0.11204	0.33917	0.13109	0.09454	0.21075	0.43408	0.38400	0.37467	0.32947	0.34987	0.30803
#05 multilearn T	0.84141	0.02831	0.00000	0.00000	0.46873	0.11661	0.00539	0.01352	0.40392	0.09848	0.02168	0.03394	0.77049	0.17370	0.69803	0.08472	0.64715	0.08334
#06 multilearn+forgets T	0.82621	0.05337	0.00000	0.00000	0.44981	0.11834	0.00639	0.01181	0.40516	0.11335	0.02558	0.03835	0.59348	0.39115	0.47465	0.32350	0.43967	0.29754
#07 multigs+multilearn T	0.72287	0.09173	0.00000	0.00000	0.41546	0.10993	0.01251	0.03567	0.35354	0.13109	0.07617	0.13654	0.68140	0.16448	0.69241	0.06171	0.64383	0.06084
#08 multigs+multilearn+forgets T	0.69074	0.15913	0.00399	0.01774	0.37808	0.11840	0.02874	0.06891	0.35203	0.13674	0.09615	0.21960	0.49700	0.36351	0.44529	0.31930	0.41427	0.29725
#09 multiprior T	0.85941	0.02847	0.00000	0.00000	0.47107	0.13325	0.00968	0.02380	0.39747	0.12624	0.03999	0.07444	0.83273	0.13556	0.65814	0.10380	0.61200	0.09845
#10 multiprior+forgets T	0.79656	0.09220	0.00000	0.00000	0.43498	0.11630	0.00807	0.01653	0.39728	0.14055	0.04354	0.07473	0.37680	0.44037	0.28812	0.34055	0.27282	0.32279
#11 multigs TE	0.72553	0.08747	0.00000	0.00000	0.41365	0.10566	0.01240	0.03489	0.34538	0.13474	0.09005	0.15156	0.67941	0.16415	0.69148	0.06283	0.64058	0.05822
#12 multigs+forgets TE	0.72138	0.09530	0.00000	0.00002	0.38056	0.10253	0.03108	0.11292	0.32703	0.13008	0.10453	0.21031	0.46693	0.37643	0.41238	0.32717	0.38307	0.30394
#13 multilearn TE	0.84059	0.02866	0.00000	0.00000	0.46701	0.11652	0.00591	0.01554	0.40367	0.09792	0.02150	0.03344	0.77040	0.17463	0.69826	0.08421	0.64740	0.08329
#14 multilearn+forgets TE	0.81900	0.06533	0.00000	0.00000	0.44595	0.12049	0.00763	0.02101	0.40022	0.12570	0.03170	0.04475	0.53493	0.43008	0.41685	0.33482	0.38913	0.31377
#15 multigs+multilearn TE	0.72593	0.08643	0.00000	0.00000	0.41367	0.10604	0.01269	0.03644	0.34474	0.13552	0.09177	0.15485	0.67941	0.16415	0.69148	0.06283	0.64058	0.05822
#16 multigs+multilearn+forgets TE	0.70175	0.12169	0.00003	0.00013	0.39053	0.11556	0.02065	0.05881	0.33489	0.12292	0.09130	0.21423	0.54371	0.35503	0.48501	0.31209	0.45115	0.28988
#17 multiprior TE	0.85941	0.02847	0.00000	0.00000	0.47107	0.13325	0.00968	0.02380	0.39747	0.12624	0.03999	0.07444	0.83273	0.13556	0.65814	0.10380	0.61200	0.09845
#18 multiprior+forgets TE	0.79656	0.09220	0.00000	0.00000	0.43498	0.11630	0.00807	0.01653	0.39728	0.14055	0.04354	0.07473	0.37680	0.44037	0.28812	0.34055	0.27282	0.32279

Table 4.22. Estimation of knowledge mastery – Average results (DKCs 01-25) per BKT model.

BKT model	Correlation												Classification					
	Formative				Summative - Midterm				Summative - Final				Formative		Summative - Midterm		Summative - Final	
	Pearson r Average	SD	p-value Average	SD	Pearson r Average	SD	p-value Average	SD	Pearson r Average	SD	p-value Average	SD	F1 Average	SD	F1 Average	SD	F1 Average	SD
#01 vanilla	0.83883	0.02751	0.00000	0.00000	0.46968	0.11314	0.00418	0.00856	0.40604	0.09694	0.02072	0.03231	0.76889	0.17352	0.69666	0.08380	0.64589	0.08309
#02 vanilla+forgets	0.81073	0.06144	0.00000	0.00000	0.43974	0.11718	0.01011	0.03128	0.39704	0.13960	0.04378	0.07438	0.45985	0.44594	0.35823	0.34733	0.33388	0.32423
Multigs models	0.71185	0.11273	0.00051	0.00641	0.39715	0.11234	0.02073	0.07004	0.34386	0.13204	0.09000	0.18292	0.58286	0.30366	0.56060	0.26806	0.52085	0.24927
Multilearn models	0.77106	0.10926	0.00050	0.00641	0.42866	0.11970	0.01249	0.03921	0.37477	0.12450	0.05698	0.13734	0.63385	0.31439	0.57525	0.26359	0.53415	0.24508
Multiprior models	0.82799	0.07512	0.00000	0.00000	0.45303	0.12636	0.00887	0.02050	0.39738	0.13359	0.04177	0.07460	0.60477	0.39764	0.47313	0.31241	0.44241	0.29275
Time-based m.	0.76798	0.11654	0.00051	0.00641	0.42542	0.12304	0.01482	0.05227	0.37534	0.12940	0.05915	0.13692	0.60837	0.33781	0.54042	0.28415	0.50289	0.26470
Time- and evals-based m.	0.77377	0.10020	0.00000	0.00005	0.42718	0.11909	0.01351	0.05096	0.36884	0.13109	0.06430	0.14098	0.61054	0.33898	0.54272	0.28307	0.50459	0.26364

Table 4.23. Estimation of knowledge mastery – Average results (DKCs 01-25) per BKT model group.

4.6 Student learning paths

Based on the learning path, which represents the sequence of binary correctness of student answers in the assessed tasks, model convergence information was also included into the BKT ranking. Student learning paths were analysed for each BKT model using the average number of answer opportunities in each examined dataset, along with the corresponding SD. The ideal learning path for each model referred to the minimum number of answer opportunities needed to achieve knowledge mastery, which included a mastery threshold of 95%.

The results regarding student learning paths are presented in Tables D1- D21 (Appendix D).

To facilitate discussion of the results, average values across DKCs 01-25 were calculated for each BKT model (Table 4.24) and model group (Table 4.25).

Table 4.24. Student learning paths – Average results (DKCs 01-25) per BKT model.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	11.54843	7.79935	11.81783	7.80907	2.28571
#02 vanilla+forgets	15.21051	4.06086	15.39569	4.08333	10.76190
#03 multigs T	8.74353	7.74191	8.97177	7.77224	1.23810
#04 multigs+forgets T	12.59674	4.07538	12.64268	4.06020	9.19048
#05 multilearn T	11.63014	7.84188	11.87795	7.84466	2.23810
#06 multilearn+forgets T	14.39143	5.27958	14.58729	5.29373	10.80952
#07 multigs+multilearn T	8.74679	7.75608	8.97179	7.78769	1.23810
#08 multigs+multilearn+forgets T	11.56967	4.79902	11.68482	4.78424	9.14286
#09 multiprior T	14.03280	6.44230	14.21627	6.43479	4.47619
#10 multiprior+forgets T	16.78059	2.95059	16.84412	2.98169	13.14286
#11 multigs TE	8.76014	7.79190	9.01825	7.84974	1.23810
#12 multigs+forgets TE	12.35836	4.72084	12.52390	4.76114	8.28571
#13 multilearn TE	11.61371	7.83221	11.86249	7.83709	2.23810
#14 multilearn+forgets TE	14.95756	4.72959	15.18154	4.74058	10.80952
#15 multigs+multilearn TE	8.74354	7.79173	9.00879	7.84631	1.23810
#16 multigs+multilearn+forgets TE	11.60050	5.44042	11.77345	5.44988	7.38095
#17 multiprior TE	14.03280	6.44230	14.21627	6.43479	4.47619
#18 multiprior+forgets TE	16.78059	2.95059	16.84412	2.98169	13.14286

Table 4.25. Student learning paths – Average results (DKCs 01-25) per BKT model group.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	11.54843	7.79935	11.81783	7.80907	2.28571
#02 vanilla+forgets	15.21051	4.06086	15.39569	4.08333	10.76190
Multigs models	10.38991	6.26466	10.57443	6.28893	4.86905
Multilearn models	11.53938	6.45668	11.74203	6.47270	5.59375
Multiprior models	15.40669	4.69645	15.53019	4.70824	8.80952
Time-based m.	12.31146	5.86084	12.47459	5.86990	6.43452
Time- and evals-based m.	12.35590	5.96245	12.55360	5.98765	6.10119

For the baseline vanilla BKT model, the average number of answer opportunities in the testing subset was 11.81783 (SD 7.80907), while in the ideal context, students achieved mastery with an average of 2.28571 answer opportunities. The multigs, multigs+multilearn and multigs+multilearn+forgets models (#03, #07, #08, #11, #15, #16) overperformed the baseline vanilla model, yielding the lowest average of 8.97177 answer opportunities (SD 7.77224 for #03) and 1.23810 opportunities in the ideal context.

In addition to the model's ability to predict student performance and estimate knowledge mastery, it was important to include the average and ideal number of answer opportunities for the BKT models. Considering the mastery threshold of 0.95, the earliest possible model convergence was crucial for drawing conclusions about student knowledge.

4.7 BKT model ranking

For each DKC, the BKT models were ranked using Composite scores derived from the normalised features. Two sets of equal feature weights were applied, depending on the availability of summative assessment data (midterm and final exams).

Composite score A incorporated the following features: student performance prediction (RMSE, AUC), knowledge mastery estimation (Correlation and F1-Formative, Correlation and F1-Midterm, Correlation and F1-Final), and adaptive answer opportunities (Average, Ideal). In this approach, ten features were assigned an arbitrary equal weight, with each receiving a value of 0.1.

Composite score B, in contrast, was based on student performance prediction (RMSE, AUC), knowledge mastery estimation (Correlation and F1-Formative), and adaptive answer opportunities (Average, Ideal). In total, six features were assigned arbitrary equal weight, with each receiving a value of 1/6 $0.1\dot{6}$.

The BKT model ranking results, based on normalised features for each DKC, are presented in Tables E1- E21 (Appendix E).

Table 4.26 provides the average values across DKCs for each BKT model, including metrics such as RMSE (SD), AUC (SD), Correlation and F1 scores (for Formative, Midterm and Final assessments). Additionally, the table includes the Average learning paths and Ideal Average learning paths. The results are visualized using heatmaps.

Based on the average results for each BKT model and considering the summative assessment data (Composite score A), two enhanced BKT models outperformed the vanilla model. The #01 vanilla model had an average rank of 5.57143 (SD 2.73551), while the #05 multilearn T model ranked at 5.14286 (SD 2.45504) and the #13 mutilearn TE model at 5.47619 (SD 3.03345).

For Composite score B, eight enhanced BKT models outperformed the vanilla BKT model (Mean 8.71429, SD 3.28261). The #11 multigs TE model achieved the best average rank of 3.80952 (SD 2.34255).

BKT model		RMSE	AUC	Correlation			F1			Aver. path	Ideal path	Composite score A	Rank A	Composite score B	Rank B
		Average	Average	Average			Average			Average	Average	Average	Average	Average	Average
#01 vanilla	Mean	0.20910	0.22639	0.81237	0.75505	0.67466	0.78489	0.85647	0.85935	0.64963	0.88634	0.67143	5.57143	0.59479	8.71429
	SD	0.16242	0.22379	0.18636	0.22976	0.27772	0.20725	0.27356	0.21238	0.18095	0.13496	0.08435	2.73551	0.08420	3.28261
	Median	0.18783	0.15843	0.84956	0.76719	0.69875	0.90000	0.96154	0.97101	0.72254	0.94444	0.67723	5.00000	0.60457	9.00000
#02 vanilla+forgets	Mean	0.49226	0.52709	0.66970	0.60147	0.60604	0.41362	0.40985	0.38344	0.28376	0.41487	0.48021	9.95238	0.46688	10.66667
	SD	0.20568	0.17526	0.23632	0.30462	0.32781	0.41907	0.43182	0.38957	0.30465	0.41531	0.21167	4.64426	0.20282	4.96016
	Median	0.48258	0.55941	0.67381	0.68259	0.70239	0.45000	0.30823	0.46970	0.18098	0.50000	0.53741	10.00000	0.45259	13.00000
#03 multigs T	Mean	0.28216	0.43293	0.30766	0.45611	0.39626	0.53988	0.82104	0.81720	0.94700	1.00000	0.60002	10.14286	0.58494	8.80952
	SD	0.22197	0.26308	0.29453	0.30043	0.36142	0.32316	0.29920	0.29868	0.06860	0.00000	0.11714	4.32364	0.10099	3.72404
	Median	0.28485	0.42034	0.26243	0.47612	0.37218	0.63656	0.94638	0.95782	0.96797	1.00000	0.62379	10.00000	0.57136	9.00000
#04 multigs+forgets T	Mean	0.59490	0.75378	0.17599	0.21654	0.36207	0.21745	0.30695	0.33379	0.56242	0.57143	0.40953	14.19048	0.47932	11.00000
	SD	0.18892	0.13105	0.22182	0.28587	0.36327	0.33204	0.40425	0.40087	0.48775	0.49487	0.16228	2.90515	0.17012	4.18614
	Median	0.56285	0.75171	0.06597	0.09514	0.24925	0.00000	0.00000	0.00000	0.96120	1.00000	0.37248	15.00000	0.53037	12.00000
#05 multilearn T	Mean	0.21924	0.23512	0.82519	0.73560	0.67883	0.78954	0.87227	0.87763	0.63980	0.88885	0.67621	5.14286	0.59962	8.00000
	SD	0.15877	0.22972	0.19768	0.22315	0.27383	0.20786	0.25375	0.19559	0.19437	0.13558	0.07956	2.45504	0.08213	2.56348
	Median	0.19815	0.17780	0.89385	0.76822	0.76121	0.90000	0.97634	0.97619	0.72043	0.94737	0.67863	4.00000	0.60433	7.00000
#06 multilearn+forgets T	Mean	0.48976	0.50235	0.75037	0.63126	0.62828	0.57586	0.56866	0.56010	0.32932	0.41871	0.54547	8.61905	0.51106	9.85714
	SD	0.19093	0.17877	0.21356	0.34995	0.31057	0.39405	0.42011	0.39021	0.27849	0.42196	0.19210	4.96153	0.17437	4.82341
	Median	0.48445	0.54484	0.82009	0.81470	0.78790	0.65185	0.67368	0.65152	0.29332	0.50000	0.60403	10.00000	0.50343	11.00000
#07 multigs+multilearn T	Mean	0.29484	0.45389	0.30904	0.45206	0.39082	0.54047	0.82148	0.81763	0.94679	1.00000	0.60270	10.04762	0.59084	8.57143
	SD	0.22239	0.25307	0.29053	0.29233	0.36720	0.32411	0.29940	0.29882	0.07116	0.00000	0.10647	4.45588	0.09477	3.61967
	Median	0.30139	0.42508	0.26294	0.47418	0.36980	0.63656	0.94638	0.95782	0.97912	1.00000	0.61594	10.00000	0.57251	9.00000
#08 multigs+multilearn+f. T	Mean	0.62044	0.71580	0.22637	0.31242	0.42539	0.26326	0.41603	0.42315	0.62051	0.57143	0.45948	13.28571	0.50297	10.28571
	SD	0.17210	0.16910	0.28613	0.29247	0.35695	0.36939	0.45008	0.44928	0.45426	0.49487	0.15581	3.07281	0.16811	4.47366
	Median	0.60600	0.71344	0.08460	0.19730	0.41898	0.00000	0.11312	0.00000	0.94797	1.00000	0.46393	13.00000	0.53189	11.00000
#09 multiprior T	Mean	0.04683	0.01075	0.95733	0.69797	0.64026	0.93489	0.69901	0.78523	0.31747	0.61007	0.56998	10.38095	0.47956	12.95238
	SD	0.11536	0.03853	0.06135	0.30625	0.32066	0.12747	0.33838	0.26902	0.22034	0.33851	0.12068	3.74453	0.10116	4.06174
	Median	0.00000	0.00000	0.99788	0.74560	0.76990	1.00000	0.80918	0.91313	0.34304	0.73684	0.58403	10.00000	0.51930	13.00000
#10 multiprior+forgets T	Mean	0.33749	0.22629	0.64513	0.58443	0.61189	0.39085	0.22901	0.29531	0.05937	0.12267	0.35024	15.04762	0.29696	16.28571
	SD	0.20726	0.14112	0.35934	0.38940	0.39450	0.46150	0.37510	0.43571	0.15881	0.30233	0.18192	4.04117	0.16144	2.96636
	Median	0.30581	0.18404	0.74240	0.78247	0.85272	0.00000	0.00000	0.00000	0.00000	0.00000	0.30645	17.00000	0.26824	17.00000
#11 multigs TE	Mean	0.63855	0.77842	0.30522	0.42804	0.35270	0.54757	0.82927	0.76361	0.93913	1.00000	0.65825	5.90476	0.70148	3.80952
	SD	0.22105	0.16113	0.29149	0.26403	0.30932	0.30764	0.27854	0.33611	0.07833	0.00000	0.10172	4.24157	0.08230	2.34255
	Median	0.65965	0.79456	0.24359	0.50605	0.23356	0.64699	0.94226	0.95522	0.95291	1.00000	0.69400	3.00000	0.68320	3.00000
#12 multigs+forgets TE	Mean	0.89882	0.98278	0.24646	0.25510	0.29609	0.27515	0.42934	0.43414	0.59436	0.61905	0.50313	10.42857	0.60277	6.38095
	SD	0.09906	0.02994	0.30825	0.29259	0.33343	0.35492	0.45222	0.43560	0.47033	0.48562	0.16071	4.34874	0.16564	4.76571
	Median	0.92421	1.00000	0.06435	0.10713	0.09913	0.00000	0.12723	0.37391	0.94479	1.00000	0.56971	10.00000	0.65436	3.00000
#13 multilearn TE	Mean	0.21494	0.22959	0.81948	0.72594	0.67514	0.78936	0.87267	0.87801	0.64153	0.88885	0.67355	5.47619	0.59729	8.38095
	SD	0.16012	0.22789	0.19829	0.21971	0.27006	0.20850	0.25392	0.19577	0.19469	0.13558	0.08055	3.03345	0.08328	3.38732
	Median	0.19820	0.17261	0.86368	0.75645	0.76150	0.90000	0.98039	0.97851	0.72254	0.94737	0.67795	5.00000	0.60430	8.00000
#14 multilearn+forgets TE	Mean	0.48560	0.53849	0.70975	0.64341	0.61580	0.51492	0.48234	0.44969	0.27403	0.42323	0.51373	8.71429	0.49100	10.19048
	SD	0.19749	0.17605	0.26523	0.34066	0.33582	0.43155	0.43859	0.41279	0.25995	0.42232	0.22929	6.04068	0.19631	5.05794
	Median	0.49672	0.58361	0.73275	0.73402	0.74774	0.58481	0.56416	0.58333	0.29701	0.50000	0.55112	10.00000	0.52535	11.00000
#15 multigs+multilearn TE	Mean	0.63896	0.77758	0.30497	0.43896	0.35061	0.54757	0.82927	0.76361	0.94125	1.00000	0.65928	5.90476	0.70172	3.80952
	SD	0.21426	0.16024	0.28930	0.27556	0.31251	0.30764	0.27854	0.33611	0.07451	0.00000	0.10215	4.54506	0.08112	2.36279
	Median	0.62773	0.79564	0.25699	0.49735	0.22822	0.64699	0.94226	0.95522	0.95291	1.00000	0.68799	4.00000	0.69054	3.00000
#16 multigs+multilearn+f. TE	Mean	0.96256	0.96370	0.22076	0.35146	0.31153	0.38960	0.55752	0.53441	0.62031	0.66667	0.55785	8.19048	0.63727	5.42857
	SD	0.06032	0.03870	0.23482	0.36565	0.31609	0.39456	0.43921	0.43371	0.41261	0.47140	0.20254	5.26099	0.20449	5.55982
	Median	1.00000	0.97245	0.15922	0.15862	0.18697	0.16800	0.84859	0.77286	0.81720	1.00000	0.66984	10.00000	0.71452	1.00000
#17 multiprior TE	Mean	0.07557	0.03525	0.95733	0.69797	0.64026	0.93489	0.69901	0.78523	0.31747	0.61007	0.57531	9.80952	0.48843	12.19048
	SD	0.11029	0.03206	0.06135	0.30625	0.32066	0.12747	0.33838	0.26902	0.22034	0.33851	0.11920	3.77169	0.09929	3.82186
	Median	0.03928	0.02877	0.99788	0.74560	0.76990	1.00000	0.80918	0.91313	0.34304	0.73684	0.58848	9.00000	0.51974	12.00000
#18 multiprior+forgets TE	Mean	0.37236	0.24939	0.64513	0.58443	0.61189	0.39085	0.22901	0.29531	0.05937	0.12267	0.35605	14.19048	0.30664	15.66667
	SD	0.20376	0.13817	0.35934	0.38940	0.39450	0.46150	0.37510	0.43571	0.15881	0.30233	0.18268	4.63546	0.16103	2.90457
	Median	0.34864	0.22597	0.74240	0.78247	0.85272	0.00000	0.00000	0.00000	0.00000	0.00000	0.30645	16.00000	0.29072	17.00000

Table 4.26. BKT model ranking – Average of results (DKCs 01-25) per BKT model.

4.8 Discussion

Hypothesis 1, regarding the feasibility of the BKT model, was addressed by considering the requirements, results, and limitations identified in the previous research tasks.

For Data collection, empirical research was conducted using an in situ quasi-experimental design within the Introductory Programming course, involving a sample of 150 students. Weekly course topics represented granulated Domain Knowledge Components (DKC) used to evaluate student knowledge through Controlled Environment (CE) formative assessments. Each test comprised 20 randomised Virtual Programming Lab (VPL)-based questions of equal difficulty and was administered at the beginning of the laboratory exercises. The in situ experimental settings were recognized as a potential limitation of this research task. Despite its limitations, the benefits of in situ experiments often outweigh the challenges, such as less controlled real-world environment and variability, especially when the goal is to provide realistic, hands-on learning experiences.

Regarding Data pre-processing, features related to the time spent on tasks and the number of code evaluations were successfully extracted for each collected CE test dataset. A classification model was developed based on time and cumulative time on task features alongside the model that additionally incorporated the number and cumulative number of student code evaluations. Although the latter classification model was tailored for the programming domain, the performance results prompted the use of both models in further research on BKT-based cognitive modelling. A potential limitation of this research task included the approximation of time per question.

Based on the pre-processed data, fitting the BKT parameters and evaluating models ensured the feasibility of the BKT model. A total of 18 BKT models were fitted, including the vanilla model, its forgetting-enhanced version, and two feature-based model groups, including time-based group and time- and evaluation-based group. The randomness was controlled using the seed parameter, the Guess and Slip parameters were bounded to 0.5, while the EM algorithm included 5 initializations. For each DKC, the BKT model that balanced student performance prediction, estimation of knowledge mastery, and timely model convergence was identified. Consequently, the evaluation results of all enhanced BKT models confirm their feasibility. The nuances in their capabilities facilitated the proposal of the Composite scores that encompassed various evaluation aspects. As a result, an additional hypothesis was proposed to establish a framework for ranking BKT models to identify the most effective model that overperformed the baseline vanilla model (Hypothesis 2).

To address Hypothesis 2, an analysis of BKT ranking results was conducted by summarizing the top-performing models along with the comment on the related not statistically significant (NSS) data (Table 4.27). This summary comprehensively evaluated model performance and highlighted which models demonstrated the most statistically significant improvements.

Table 4.27. The best performing BKT models per DKC - Composite scores A and B.

DKC	Composite score A		Composite score B
	The best model ranking	**Statistical significance**	**The best model ranking**
DKC 01	#06 multilearn+forgets T	NSS (Estimation-Correlation-Final exam)	#06 multilearn+forgets T
DKC 02	#06 multilearn+forgets T		#14 multilearn+forgets TE
DKC 03	#05 multilearn T	NSS (Prediction-RMSE)	#11 multigs TE
DKC 04	#14 multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 05	#15 multigs+multilearn TE		#06 multilearn+forgets T
DKC 06	#15 multigs+multilearn TE		#11 multigs TE
DKC 07	#12 multigs+forgets TE		#02 vanilla+forgets
DKC 08	#13 multilearn TE	NSS (Prediction-RMSE)	#07 multigs+multilearn T
DKC 09	#11 multigs TE		#15 multigs+multilearn TE
DKC 10	#15 multigs+multilearn TE		#15 multigs+multilearn TE
DKC 11	#14 multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 12	#17 multiprior TE		#16 multigs+multilearn+forgets TE
DKC 13	#15 multigs+multilearn TE		#16 multigs+multilearn+forgets TE
DKC 14	#16 multigs+multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 15	#14 multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 16	#05 multilearn T	NSS (Prediction-RMSE, Estimation-Correlation-Final exam)	#16 multigs+multilearn+forgets TE
DKC 17	#15 multigs+multilearn TE		#11 multigs TE
DKC 19	#16 multigs+multilearn+forgets TE	NSS (Estimation-Correlation-Final exam)	#16 multigs+multilearn+forgets TE
DKC 21	#16 multigs+multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 23	#16 multigs+multilearn+forgets TE		#16 multigs+multilearn+forgets TE
DKC 25	#14 multilearn+forgets TE		#16 multigs+multilearn+forgets TE

For Composite score A, which accounted for both formative and summative assessments, the enhanced BKT models (#02-#18) outperformed the baseline vanilla model in all DKCs. Among the enhanced models, the multilearn+forgets TE (#14), the multigs+multilearn TE (#15) and the multigs+multilearn+forgets TE (#16) models most frequently achieved the best performance, with 13 out of 21 DKCs. The statistical significance for differences with the baseline vanilla model was not established in five DKCs. Specifically, DKC 01, DKC 16 and DKC 19 lacked significance in correlation results with the Final exam, while DKC 03, DKC 08 and DKC 16 for the prediction results of the RMSE metric.

For Composite score B, which focused solely on formative results, the enhanced models (#02-#18) consistently outperformed the baseline vanilla model across all DKCs. The multigs+multilearn+forgets (#16) BKT model was the most frequent top performer, appearing as best in 11 out of 21 DKCs. Statistical significance was established across all DKCs.

Although the vanilla BKT model demonstrated strong average performance, it was consistently surpassed by the enhanced BKT models across all DKCs. The variation in model performance, particularly the repeated prominence of the multigs+multilearn+forgets (#16) BKT model in the later DKCs, indicated a more robust fit for more complex learning tasks.

Overall, the enhanced BKT models provided a more nuanced understanding of student learning processes, making them the more effective choice for modelling student knowledge.

5 DEMONSTRATION OF THE BKT-BASED ADAPTIVE FORMATIVE ASSESSMENT

To address the hypotheses, a data mining approach encompassing the research tasks of Data collection, Data pre-processing, BKT parameter fitting and BKT model evaluation was proposed. Furthermore, BKT-based adaptive formative assessment was implemented in the widely used Moodle LMS, which does not provide adaptive assessment. Although the assessment aimed to include any type of question, VPL-based questions were set as research prerequisites. In general, the adaptive assessment implementation integrated the Moodle VPL activity module, the VPL Question type, and the pyBKT library for cognitive modelling. Two types of Moodle modules were developed, including quiz report and activity module types.

Regarding the Data collection task, the BKT Quiz Report (BKT-QR) prototype module builds on the typical Moodle quiz (Initial assessment) and provides the report for each quiz question (Initial BKT report). The BKT Quiz (BKT-Q) prototype module uses the Initial BKT report and provides the fitting and evaluation of BKT models. It builds on the BKT API, which implements the functionalities of Data pre-processing, BKT parameter fitting, and model evaluation tasks. Based on the selected BKT model, the Moodle LMS provides adaptive assessment. The relationships between the research tasks, the BKT API, and the prototype modules are presented in Figure 5.1.

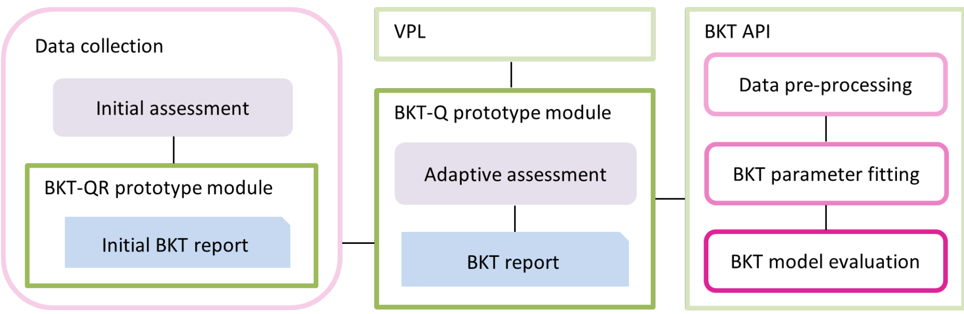

Figure 5.1. Research tasks and Moodle LMS prototype modules.

In the furter subsections, the prototype modules and the BKT API are described. For the BKT-QR prototype module, the use cases, database tables used to extract data, the prototype structure, and related classes and methods are described (Section 5.1). The BKT API's structure, data handling, and related endpoints for data retrieval and submission are presented in Section 5.2. For the BKT-Q module, the dependencies and the high-level prototype architecture are presented (Section 5.3). While the dependencies include the standard Moodle quiz engine and VPL modules, the prototype architecture illustrates interactions between a user, Moodle LMS, BKT-Q prototype module and BKT API.

Also, in situ experimental guidelines are presented, including the Self-practice and adaptive assessments (Section 5.4.1), and the usability study (Section 5.4.2).

5.1 BKT Quiz Report prototype module

The BKT-QR prototype module generates a report for a typical Moodle quiz with questions presented on a single page and sequential navigation. Along with standard question data, it presents the time spent on a task and the number of code evaluations for the VPL-based questions. The resulting Initial BKT report is primarily designed for use by the BKT-Q prototype module. The development environment included the Moodle LMS (v4.1.5+) and the XAMPP package consisting of Apache HTTP Server (v2.4.58), MariaDB database (v10.4.32) and PHP (v8.0.30).

The use case scenario for the BKT-QR prototype module refers to a teacher who generates a report and the sub-scenario of setting the report option to include not-reached answer opportunities. Once presented, a teacher can download the report in a selected format (e.g., Coma separated values, Microsoft Excel, etc.).

The server-sided reporting relies on the standard and extensive Moodle LMS database [123]. The BKT-QR prototype module bridges the unique student quiz attempt and related question data. Figure 5.2 shows the data model utilised by the BKT-QR prototype module, referring to the core tables of course, quiz, user and question. More specific tables on the quiz and question attempts provide details such as time spent on the task and the number of code evaluations.

Regarding the time aspect, the `timecreated` attribute of the `question_attempt_steps` table is leveraged, as well as the question's complete state attribute or the submitted student answer. The time spent on a task is calculated as a difference between the previous and current question timestamps.

The `question_attempt_step_data` table provides the number of student code evaluations. The VPL question evaluation results in the `_evaldata name` attribute and the related JSON-formatted value containing the number of evaluations (`nevaluations` in Figure 5.3). The number of evaluations per question is calculated based on the iteration over question steps. Time and evaluation features are set to -1 for not-presented or not-reached questions.

The BKT-QR prototype module follows the standard Moodle architecture, including the default classes' extensions, the default file and folder structure and the target `mod/quiz/report` path. It consists of the components presented in Table 5.1.

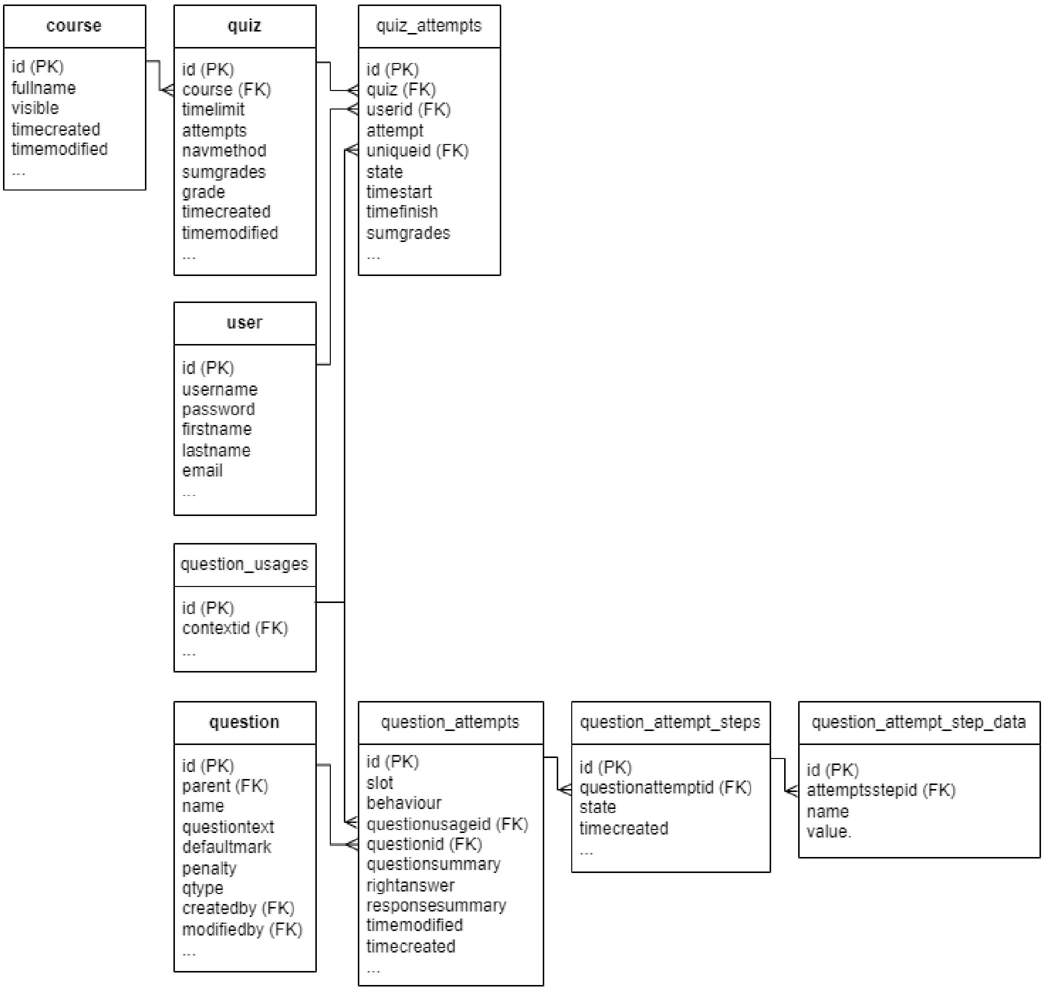

Figure 5.2. Moodle LMS database tables utilised by the BKT-QR prototype module.

```
{
  "compilation": "",
  "evaluation": "-Test 1: test1 (-100.000
    )\nRješenje nije točno\n\n-Summary of tests\n
    >+------------------------------+\n>|  1  test
    run/ 0 tests passed |\n
    >+------------------------------+\n\n",
  "execution": "",
  "grade": "Proposed grade: 0 / 100",
  "nevaluations": 6,
  "freeevaluations": "0",
  "reductionbyevaluation": "0"
}
```

Figure 5.3. The example of the evaluation data in the question_attempt_step_data table.

Table 5.1. The BKT-QR prototype module components.

Component	Description
settings.php	Mandatory file that defines the administrative settings
version.php	Mandatory file that manages the version information
lang/en/quiz_bkt.php	Mandatory file that manages the language support
bkt_options.php	Additional file that manages options for the Initial BKT report The quiz_bkt_options class extends the default mod_quiz_attempts_report_options class and adds new functionality to handle multiple quiz attempts and set up teacher preferences.
bkt_form.php	Additional file that manages form-related BKT functionalities The quiz_bkt_form class extends the default mod_quiz_attempts_report_form class and provides new functionality that allows teachers to include not-reached answer opportunities.
question_bkt_table.php	Additional file that handles the display and processing of question data in the Initial BKT report The quiz_question_bkt_table class extends the default quiz_attempts_report_table class and adds new functionalities such as tracking time on a task, counting student code evaluations, handling not-reached answer opportunities and building the report table.
report.php	Mandatory file that handles the generation and display of the Initial BKT report The quiz_bkt_report class extends the default quiz_attempts_report class and integrates the custom functionalities into Moodle's quiz reporting framework, including displaying standard question data, adding the response times and evaluation counts and handling teacher preferences for display and download formats.

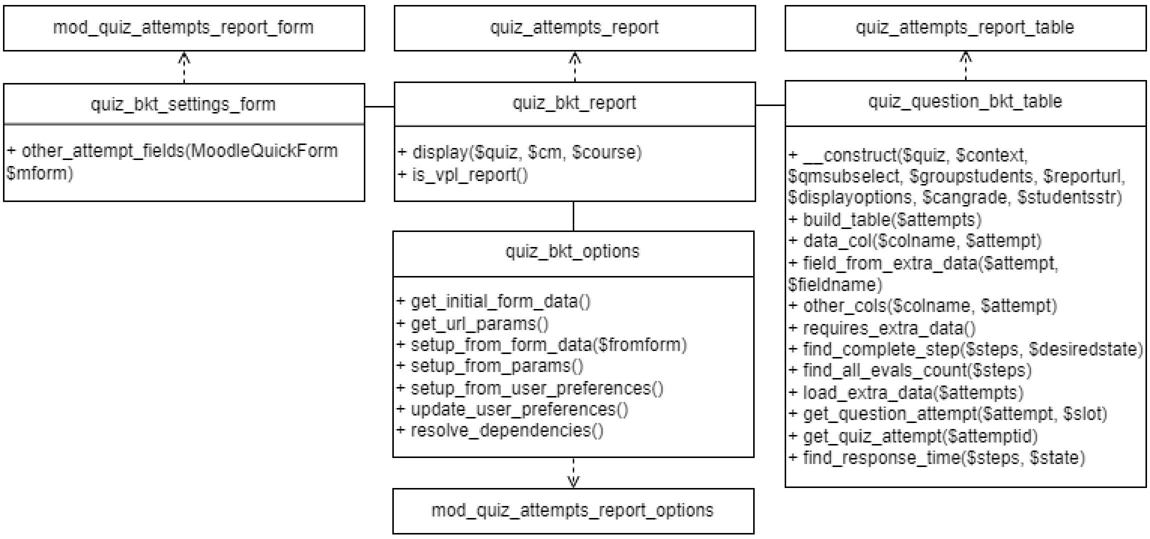

Figure 5.4. Class diagram of the BKT-QR prototype module.

Figure 5.4 presents the Unified Modelling Language (UML) class diagram of the BKT-QR prototype module. All classes extending standard Moodle classes are presented as dependency relationships. The `quiz_bkt_report` class uses the `quiz_bkt_options` class to manage the report's settings and options, the `quiz_bkt_settings_form` class to present and handle form elements and the `quiz_question_bkt_table` class to manage how the table is displayed (all presented as association relationships). The `quiz_bkt_settings_form` class allows teachers to include not-reached answer opportunities in the report. The `quiz_bkt_options` class independently encapsulates the configurable settings for the Initial BKT report, while the `quiz_question_bkt_table` renders table data.

The UML sequence diagram is presented in Figure 5.5. The sequence begins with the `display()` method in `quiz_bkt_report`, which manages the entire report generation process. The teacher requests the report, which triggers the `init()`

method. During this initialization, the `quiz_bkt_options` object is created. Next, the `process_settings_from_form()` method processes the settings submitted via the form, applying them to the options object. Afterward, `setup_from_form_data()` is called to refine the options further, ensuring the settings are accurately initialized. The `set_data()` method then updates the form with the prepared options and configuration for rendering. Finally, `get_initial_form_data()` retrieves the initial settings, prepopulating the form fields with the default or previously saved values. These methods, as part of the `quiz_bkt_report` class, ensure the form is correctly processed, validated, and preloaded.

Once the form is set up, several internal methods within `quiz_question_bkt_table` are called to process the data and manage the table columns. First, the constructor initializes the table with necessary parameters, such as quiz details, context, and display options. Then, `build_table($attempts)` is called to begin generating the table, processing the attempt data and preparing it for rendering. The `data_col($colname, $attempt)` method formats and returns the data for each column based on the attempt details. If additional data is needed for specific columns, `field_from_extra_data($attempt, $fieldname)` retrieves the required field data from extra data linked to the attempt. For custom columns like responses or right answers, `other_cols($colname, $attempt)` calls `data_col()` to handle the data for those columns.

As the table is populated, `requires_extra_data()` is called to check whether extra data, such as response times or evaluations, is needed for the table. If extra data is required, `load_extra_data($attempts)` is invoked to interact with the Moodle database and fetch the supplementary information. Methods like `find_complete_step($steps, $desiredstate)` and `find_response_time($steps, $state)` are used during this process to calculate response times and identify the completion status of each question attempt. Additionally, `get_question_attempt($attempt, $slot)` retrieves specific question attempt data, while `get_quiz_attempt($attemptid)` is used to fetch the quiz attempt data. Once all the data is processed and formatted, the completed report is ready to be presented to the teacher, concluding the sequence.

The resulting Initial BKT report provides the following information: Quiz ID, Student ID, First name / Last name, Answer opportunity, Question ID, Grade, Quiz max grade, Quiz grade %, Student performance, Time taken (sec), Cumulative time, Response evals, Cumulative evals, Student answer and Right answer.

5.2 BKT API

The BKT Application Programming Interface (API) links the pyBKT Python library for cognitive modelling (described in 3.3.) and the Moodle LMS. Its primary functions include reading Moodle quiz data (the Initial BKT report), fitting and evaluating the proposed BKT models, and updating students' knowledge mastery levels. The BKT API was developed

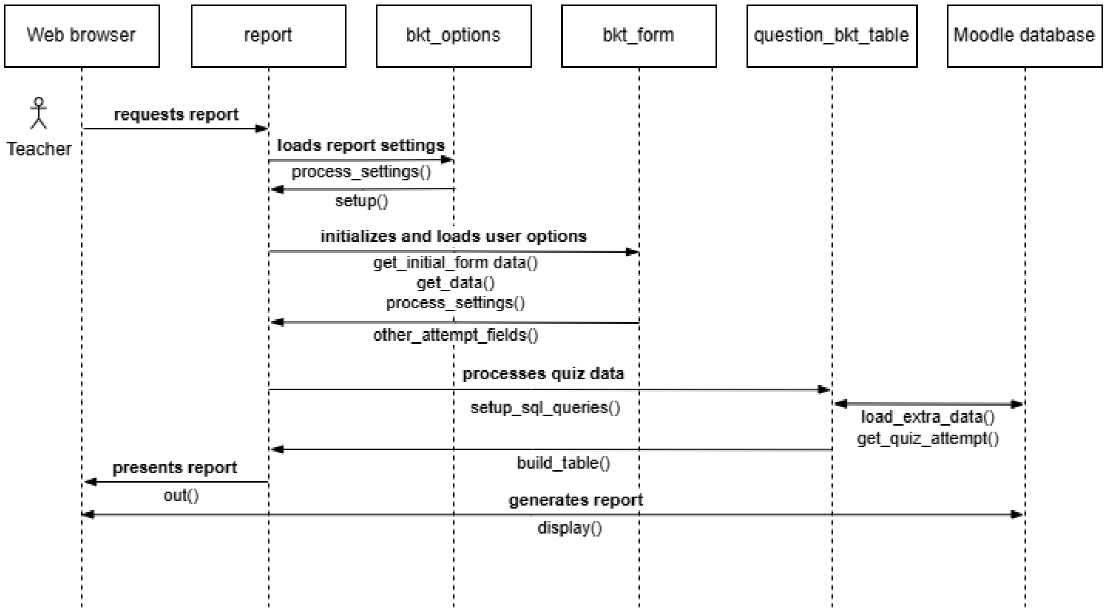

Figure 5.5. Sequence diagram of the BKT-QR prototype module.

Table 5.2. The BKT API components.

Component	Description
quiz_processor.py	Flask BKT API file
requirements.txt	File including Python dependencies, used by the Dockerfile
Dockerfile	Docker file type used to containerise the BKT API
storage	Folder used to store the resulting data
readme.md	Documentation file

using the Flask Python web framework (version 2.3.2) and Docker service (version 7.1.0) to facilitate containerisation, thereby ensuring consistent environmental settings. Libraries integrated within this environment include pandas (version 1.4.1), numpy (version 2.0.1), scikit-learn (version 1.5.1), scipy (version 1.14.0) and pyBKT (version 1.4.1). The API's functionality was tested with the Postman tool.

The specific components of the BKT API are presented in Table 5.2.

The BKT API uses Python's Pickle module to manage data processing, enabling serialization and storage of BKT and classification models in pickle files (.pkl). Upon retrieval, data is deserialised and returned in a structured JSON object format. The data components stored within the API's storage folder include:

- classification models: Decision tree-based models
- classification model evaluation metrics: AUC, F1 score, Precision, Recall, Accuracy
- BKT models: pyBKT models defined by Prior, Guess, Slip, Learn and Forgets parameter probabilities
- BKT model evaluation metrics

Table 5.3. BKT API endpoints.

Endpoint path	**Description**	**Parameters (GET) Request data (POST)**	**Response**
/train_models_from_data (POST request)	BKT model fitting and evaluation	data, bktquizid	Returns a JSON array of the classification model fitting and evaluation results and the BKT model fitting and evaluation results
/calculate_mastery (POST request)	Calculates the current knowledge mastery probability based on classification results and using the pyBKT Roster	bktquizid, trainedquizid, student_id, correct, modelid, template_id	Returns the current classification data and the student's knowledge mastery probability
/get_bkt_model (GET request)	Retrieves data for the selected BKT model	modelid, bktquizid	Returns a JSON object containing the classification models, BKT model parameters, cross-validation results, evaluation metrics and answer opportunity metric
/save_model_for_quiz (POST request)	Saves the BKT model for a specific quiz	courseid, modelid, bktquizid, trainedquizid	Returns a success message with the save key of the BKT model
/remove_model (POST request)	Removes a specific BKT model	quiz_id. courseid. modelid	Returns a success message
/get_models (GET request)	Retrieves a list of all available BKT models	-	Returns a JSON array of BKT models

- student performance prediction metrics: cross-validation metrics based on the training subset (RMSE, AUC, Accuracy) and evaluation metrics based on the testing subset (RMSE, AUC, Accuracy)

- knowledge mastery estimation metrics: correlation metrics of Pearson r and p-value and prediction metric of F1 score

- answer opportunity metrics: the average and ideal number of answer opportunities.

Regarding knowledge mastery estimation, it is important to note that the BKT API does not consider summative assessment (midterm or final exam) data. Since this data may not be directly related to the Moodle database, including it could be explored as part of future research.

The BKT API is constructed using a standardized Flask framework, which includes key functionalities such as data preprocessing, model fitting and evaluation, knowledge mastery calculation, file path generation for data storage, and helper methods for data storage and retrieval.

The BKT API endpoints, listed in Table 5.3, define specific pathways for handling requests related to data retrieval and submission.

5.3 BKT Quiz prototype module

The BKT-Q prototype module utilizes the Initial BKT report from the BKT-QR module to enhance the adaptive assessment. It allows for fitting and evaluating BKT models while continuously updating student knowledge mastery by integrating the pyBKT cognitive modelling library of the BKT API.

The BKT-Q prototype module leverages Moodle's standard question engine, the complex underlying system responsible for managing questions, student answers, grading, and feedback. One of its central components, Question Usage By Activity (QUBA), manages question instances in activities by tracking their usage, states, and outcomes. When a student interacts with a question, QUBA creates an attempt, logs each answer in a sequence of steps, and records every transition from an initial answer to final grading. More specifically, the `question_attempt` class consists of a list of `question_attempt_steps` and is responsible for the complete history of question states. Although limitedly documented, the question engine was described using sequence UML diagrams for displaying a quiz page and processing student answers (Figure 5.6) [4].

When a student accesses a quiz page in Moodle (e.g. by visiting `modquizattempt.php?id=123`), the system initiates the data retrieval process by executing `load_usage(123)` to access the quiz attempt data. The `question_engine` component then creates a new quiz attempt instance($quba with ID 123) and begins the rendering process by invoking `render_question(1, $qopt)`, where `$qopt` specifies display options for each question. During rendering, the engine generates various output components such as `$qoutput` and `$qimoutput`, responsible for assembling different parts of each question's content. These output components collaborate to construct the HTML structure, integrating text, images, and interactive elements. The final HTML content is then sent to the student's browser, enabling them to view and interact with the quiz questions directly.

When a student submits quiz answers via a POST request to `/mod/quiz/processattempt.php`, Moodle's backend processes the submission, including details such as the `quiz attempt=123` and individual answers (e.g. q123_1_answer=frog). Initially, load_usage(123) retrieves the relevant quiz attempt data, and `process_all_actions()` manages the processing of each submitted answer. For each question, Moodle calls `set_submitted_data()` to record the response, `get_expected_data()` to retrieve the required data, and validates the input. The interaction component evaluates each answer using `process_action()` and assigns a grade with `set_fraction()` based on response accuracy. Each grading instance is temporarily stored in pendingstep, documenting each part of the submission. After processing all responses, Moodle calls `save_usage($quba)` to store the updated attempt data, then redirects the student to a page displaying their updated quiz status or feedback.

Moreover, the VPL activity module is a prerequisite for the proposed BKT-based assessment, functioning as an intermediary that enables clients to submit and monitor programming tasks. These tasks are executed on specialized servers that process the code and provide feedback to both Moodle and the client (Figure 5.7).

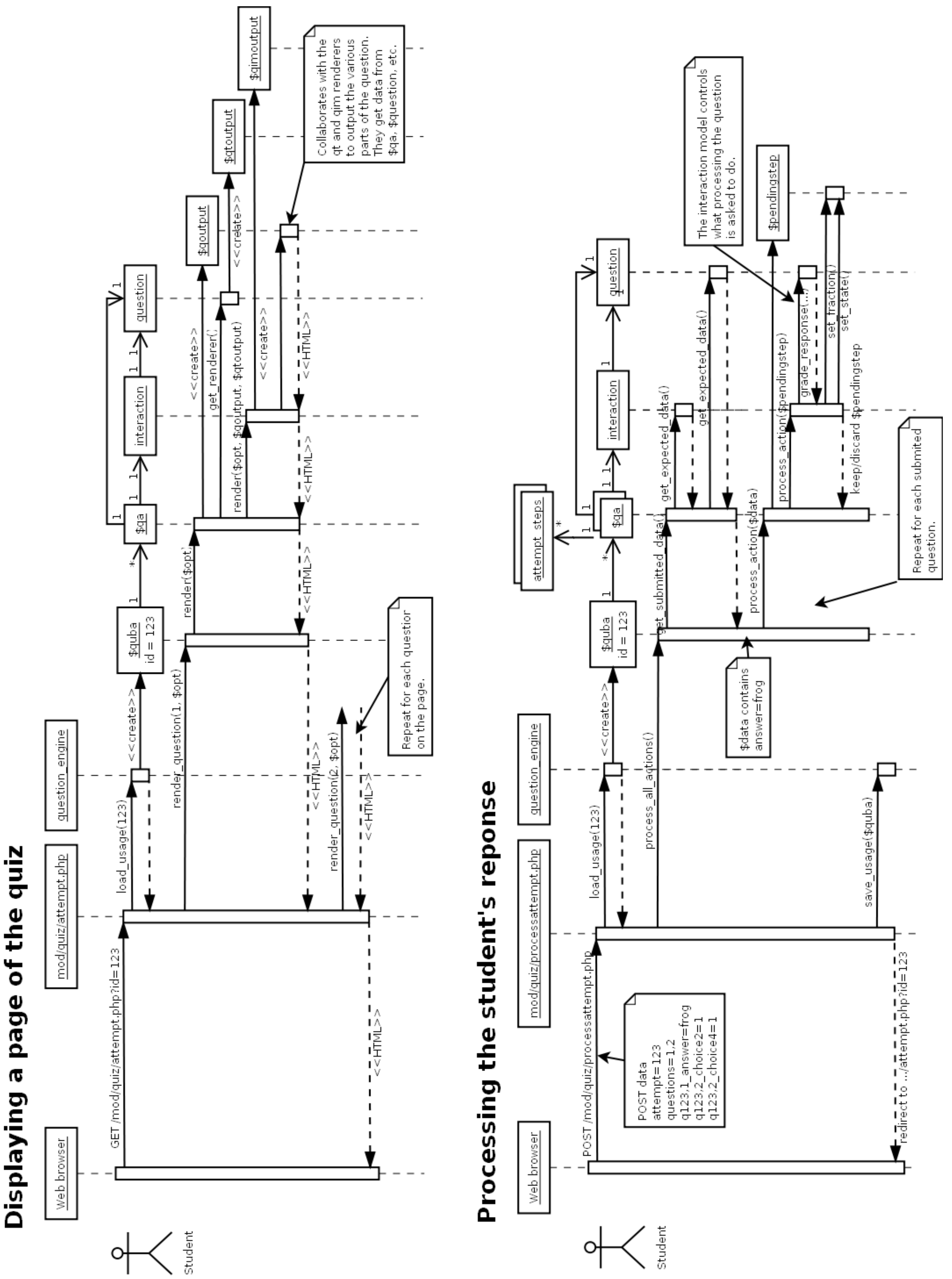

Figure 5.6. Moodle question engine: display a quiz page and processing student answers [4].

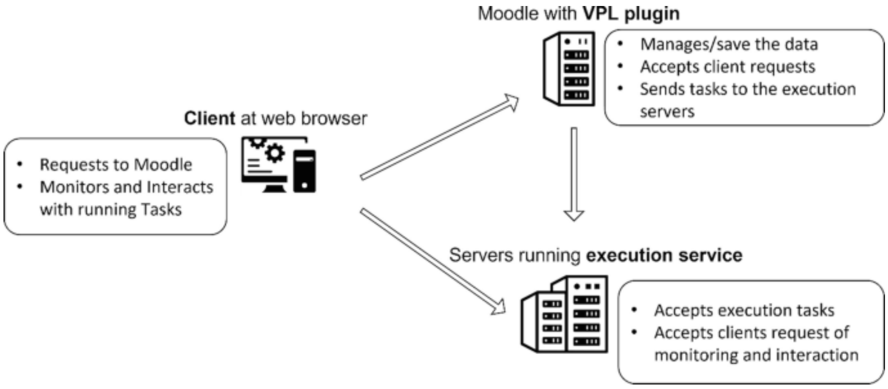

Figure 5.7. Moodle VPL activity module [5].

The VPL Question module enables continuous integration of the VPL activity module within Moodle quizzes (Figure 5.8). In this workflow, when a student submits code (e.g. `print("Hello World")`), it is inserted into a pre-configured file template set up by the teacher, which includes a `ANSWER` placeholder. This placeholder is dynamically replaced with the student's code submission, resulting in a fully executable file. Additionally, supplementary files and scripts are integrated to facilitate testing and configure the required environment. The VPL then executes these files within a controlled environment, allowing for automated code evaluation and providing immediate feedback to the student.

The BKT-Q prototype module is built on Moodle's standard activity module framework, integrating the CAT (Computer-Adaptive Testing) implementation for the Moodle activity module [124]. This module allows teachers to design assessments that evaluate student abilities using the CAT algorithm [125], as discussed in [126]. Initially created through a collaborative effort between Middlebury College and Remote Learner, the original repository was archived in 2022 [127]. Due to its wide application, with over 677 active sites as of May 2024, further development has continued through other forks of the project [127]. Moreover, the CAT activity module has been recognized and used as a base for new Moodle plugins focused on adaptive testing [128]. The development environment for the BKT-Q prototype included Moodle LMS (v4.1.5+) and the XAMPP package, consisting of Apache HTTP Server (v2.4.58), MariaDB database (v10.4.32) and PHP (v8.0.30).

The use case scenario for the BKT-Q prototype module involves a teacher who configures an adaptive assessment by specifying a mastery threshold probability (e.g. a default value of 0.95) and a question pool, uploading a CSV file to fit and evaluate the proposed BKT models, and selecting the BKT model to be used for the adaptive assessment. Based on student quiz attempts, a teacher downloads a BKT report containing knowledge mastery probabilities for each quiz question. From the student's perspective, the use case scenario of the adaptive assessment begins with the quiz attempt, dynamically updating the knowledge mastery probability, and concluding the assessment when the probability exceeds the set threshold. The architecture of the BKT-Q prototype module, illustrating the interactions

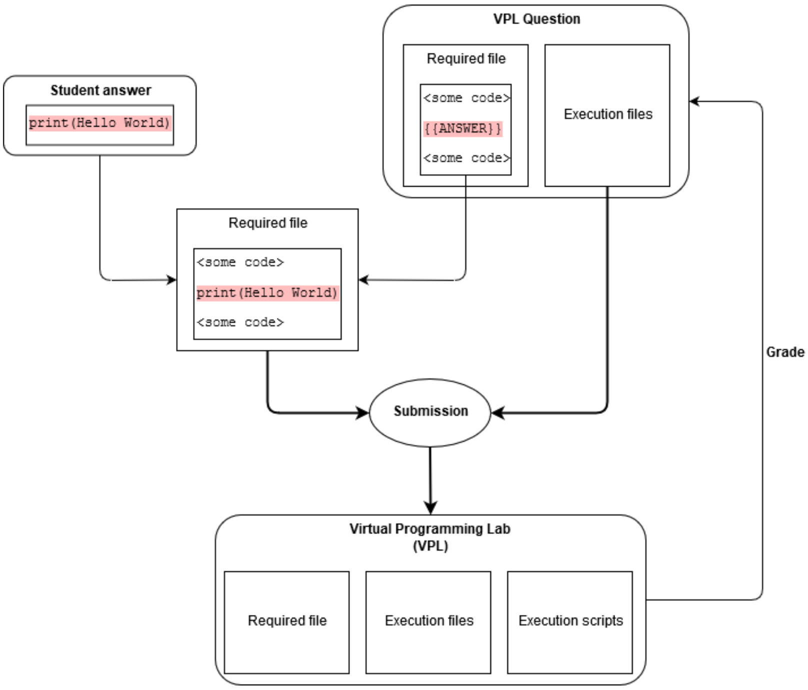

Figure 5.8. Moodle VPL Question module [6].

between the user, Moodle, the BKT-Q prototype module and the BKT API, is presented in Figure 5.9. The user, representing a teacher or student within the Moodle LMS environment, interacts with the BKT-Q prototype module. The API Interface acts as an intermediary, forwarding requests between the Moodle BKT-Q prototype module and the BKT API. The BKT API processes these requests and stores relevant data, such as classification results and BKT model parameters, in its internal `storage/` folder.

In BKT-based assessment, both client-side and server-side components collaborate to provide an adaptive quiz experience. Users interact with the quizzes on the client side, which comprises the User Interface within Moodle, submit their answers, and view feedback. User actions on the client side trigger `HTTP POST` and `GET` requests sent to the server. The server side includes Moodle's backend, the BKT-Q prototype module and the BKT API. The BKT-Q prototype module manages quiz attempts, handles HTTP requests directed to the BKT API and uses returned knowledge mastery data to update the student model and adapt the quiz dynamically. The BKT API, which runs on a separate server, handles specific requests that update student models. The Database Layer in Moodle stores user data related to quiz attempts, while the BKT API accesses its storage folder containing classification and BKT model data. This architecture ensures that user actions on the client side are captured while the server side processes the data and adjusts quiz questions according to the student's evolving mastery, facilitating a personalized and adaptive learning experience.

The BKT-Q prototype module leverages three additional database tables and introduces new fields to support the BKT-based assessment functionality. The main configuration table,

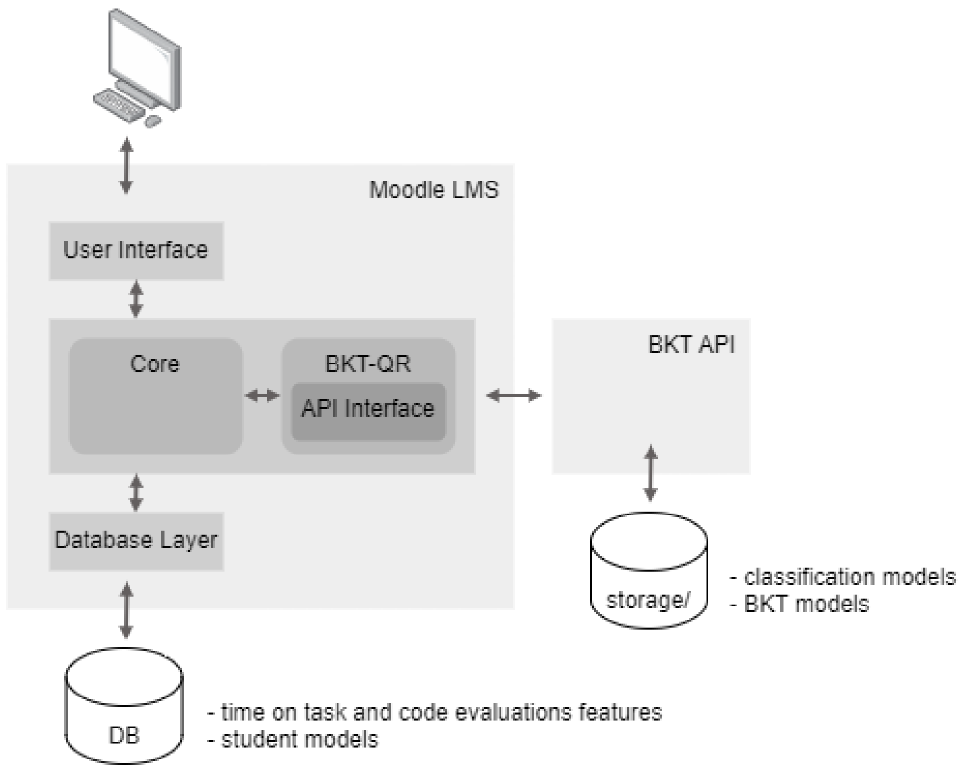

Figure 5.9. BKT-based assessment architecture.

`bktquiz` table, stores settings for each adaptive quiz instance, including the quiz name, the allowed number of attempts, mastery requirements and selected BKT model. The `bktquiz_question` table manages quiz-to-question associations, linking each quiz to relevant question categories within the question bank for adaptive selection. Meanwhile, the `bktquiz_attempt` table logs each user's quiz attempt, tracking the student's progress, including the number of questions attempted, the current knowledge mastery probability and any conditions prompting quiz termination. This table contains specific fields for BKT data, such as `masterydata`, which stores detailed information on mastery progression across attempts. The key fields introduced and used by the BKT-Q prototype module are presented in Table 5.4.

In the initial use case scenario, a teacher creates an adaptive assessment as a new activity (Figure 5.10) and sets the mastery threshold and the question pool for adaptive selection (Figure 5.11). Then, the teacher uploads a CSV file to fit and evaluate the BKT models (Figure 5.12). Following an analysis of the model information presented, the teacher selects the appropriate BKT model for the adaptive assessment (Figure 5.13). Upon this selection, the system displays the BKT parameters (Figure 5.14) along with classification model data (Figure 5.15). Once the assessment is underway, the teacher has the option to download a detailed report of the assessment results, which includes the probability of knowledge mastery for each individual question (Figure 5.16).

Table 5.4. Key fields in the BKT-Q prototype module database tables.

Table	**Field**	**Type**	**Description**
bktquiz	id	Integer	Primary key, unique identifier for each BKT-enabled quiz instance
	course	Integer	Foreign key linking the quiz to a specific course
	name	String	Name of the quiz instance
	attempts	Integer	Number of attempts allowed for the quiz
	masteryrequired	Float	Required mastery level to complete the quiz
	modeldata	JSON	Data associated with the BKT model for adaptive learning
	selectedmodelid	Integer	Identifier for the specific BKT model used in the quiz
bktquiz_question	id	Integer	Primary key, unique identifier for each question-category association
	instance	Integer	Foreign key linking to bktquiz, representing the quiz instance
	questioncategory	Integer	Foreign key linking to question_categories, specifying the questions used in the quiz
bktquiz_attempt	id	Integer	Primary key, unique identifier for each quiz attempt
	userid	Integer	Foreign key linking to user, identifying the student making the attempt
	questionsattempted	Integer	Total number of questions attempted by the user in this session
	masterydata	JSON	JSON-formatted field for tracking responses, mastery progress, and updates through each attempt
	attemptstopcriteria	String	Indicates the reason for stopping the attempt, such as mastery achieved or max questions reached

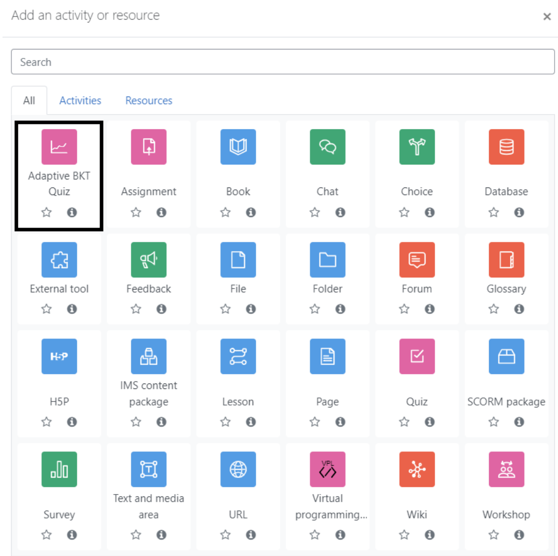

Figure 5.10. Adaptive BKT Quiz.

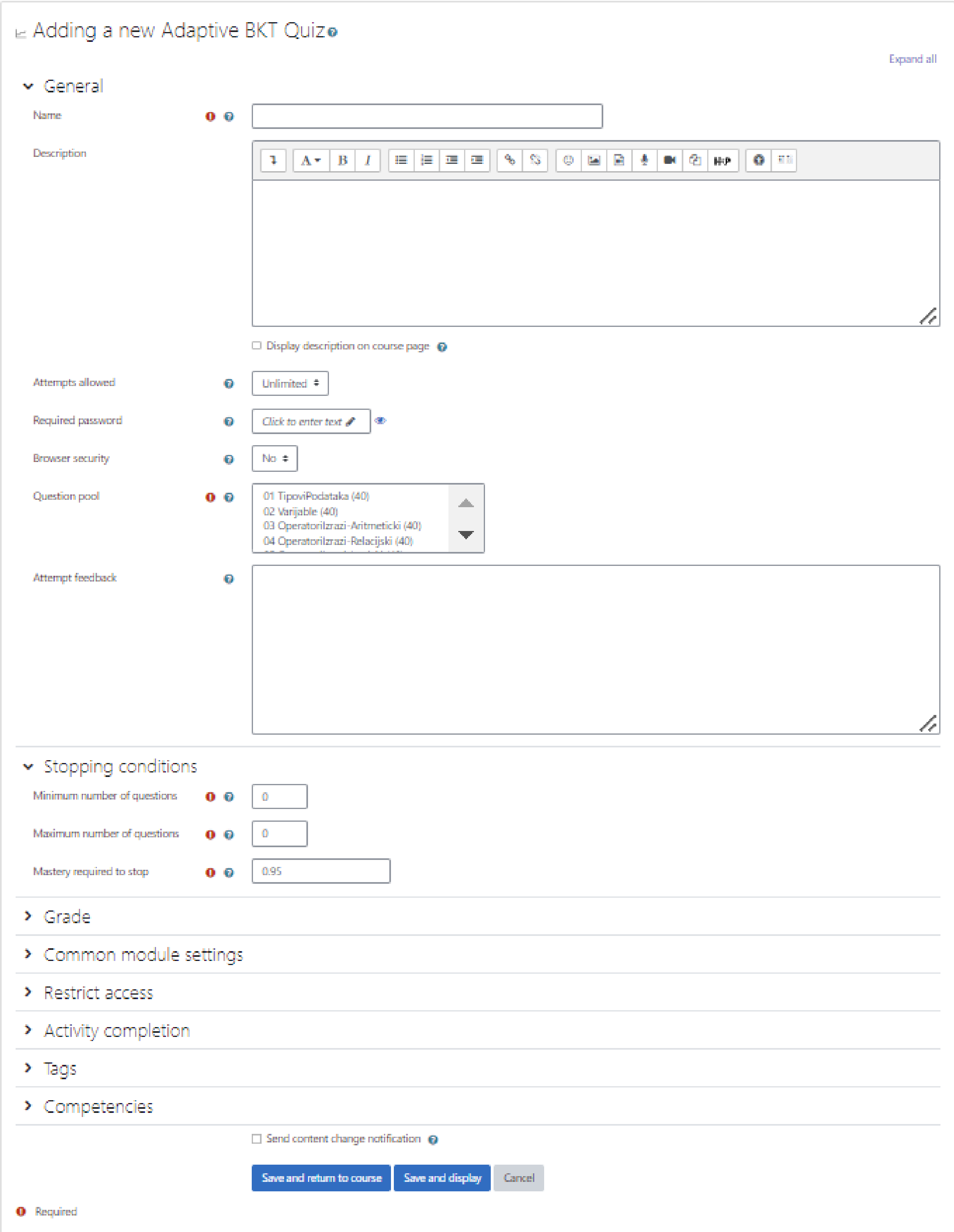

Figure 5.11. Stopping conditions – Mastery required to stop.

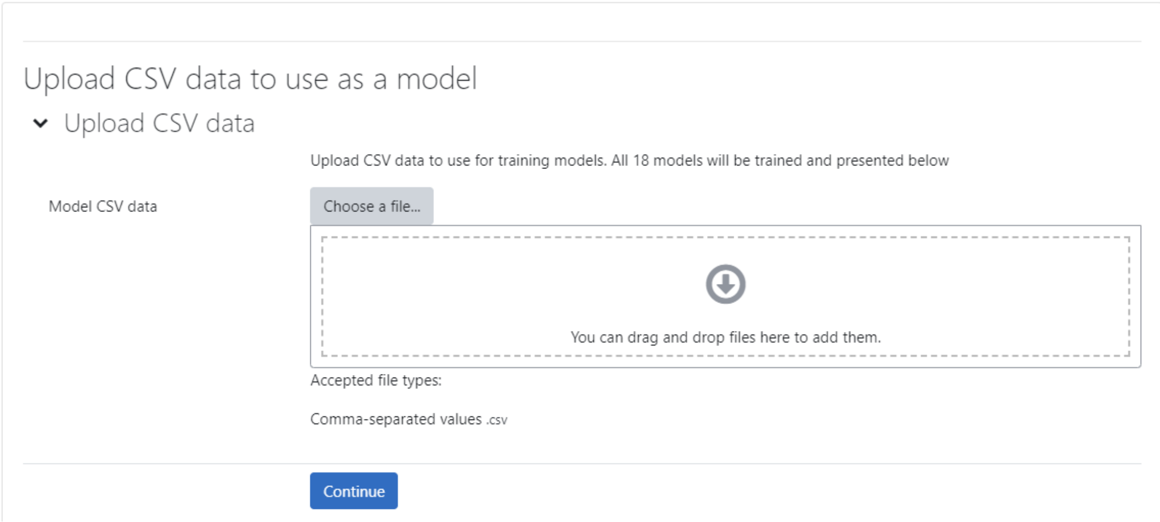

Figure 5.12. Upload CSV data.

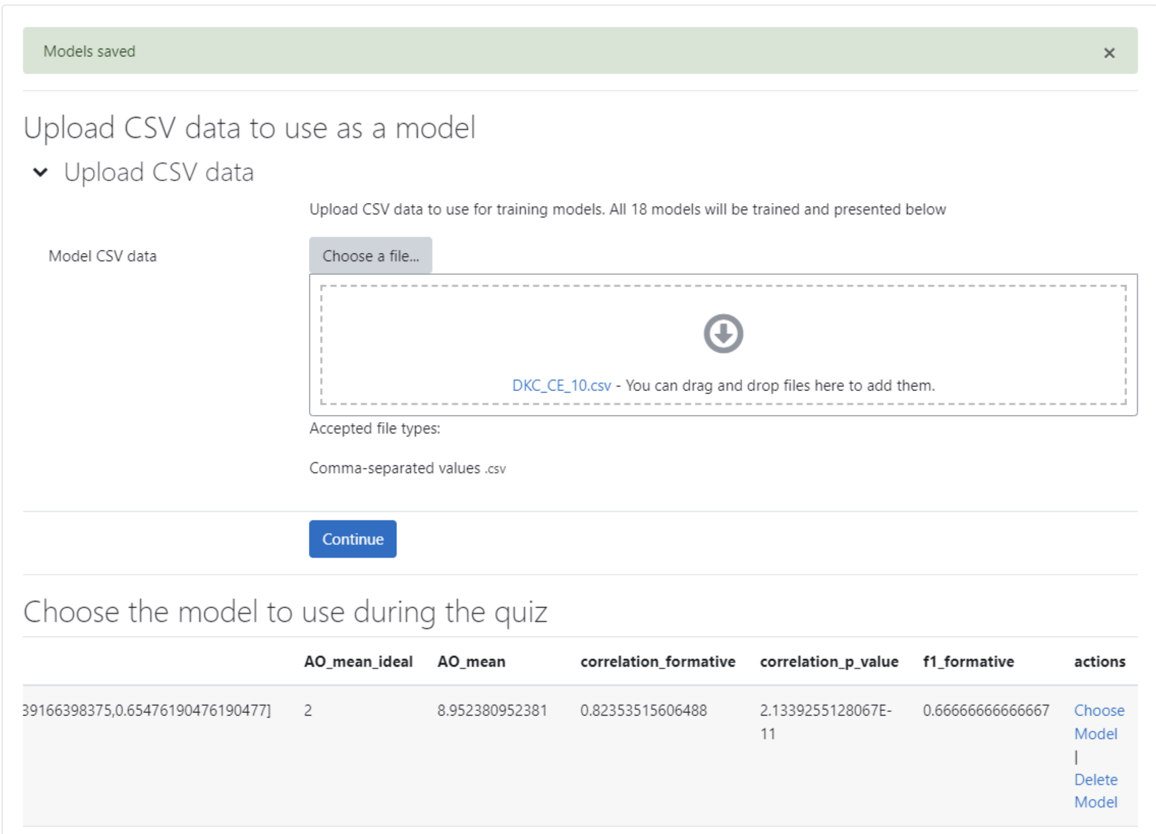

Figure 5.13. Choose BKT model.

Selected Model

modelid	source	actions
BKT_01	csv	Remove Model

Probabilities

param	class	value
prior	default	0.4537644388
learns	default	0.0143431534
guesses	default	0.0243749633
slips	default	0.4483132075
forgets	default	0

Figure 5.14. Selected BKT model data.

Tree Visualizations

clf_time_export

```
|--- time <= 13.50
|   |--- weights: [319.00, 0.00] class: 0
|--- time >  13.50
|   |--- time <= 55.50
|   |   |--- cumulative_time <= 135.00
|   |   |   |--- time <= 30.50
|   |   |   |   |--- weights: [12.00, 0.00] class: 0
|   |   |   |--- time >  30.50
|   |   |   |   |--- cumulative_time <= 47.50
|   |   |   |   |   |--- weights: [4.00, 1.00] class: 0
|   |   |   |   |--- cumulative_time >  47.50
|   |   |   |   |   |--- weights: [2.00, 4.00] class: 1
|   |   |--- cumulative_time >  135.00
|   |   |   |--- time <= 15.50
|   |   |   |   |--- cumulative_time <= 832.50
|   |   |   |   |   |--- weights: [4.00, 0.00] class: 0
|   |   |   |   |--- cumulative_time >  832.50
|   |   |   |   |   |--- weights: [1.00, 1.00] class: 0
|   |   |   |--- time >  15.50
|   |   |   |   |--- cumulative_time <= 1060.00
|   |   |   |   |   |--- weights: [45.00, 111.00] class: 1
|   |   |   |   |--- cumulative_time >  1060.00
|   |   |   |   |   |--- weights: [13.00, 10.00] class: 0
|   |--- time >  55.50
|   |   |--- time <= 140.50
|   |   |   |--- cumulative_time <= 224.50
|   |   |   |   |--- time <= 109.00
|   |   |   |   |   |--- weights: [24.00, 1.00] class: 0
|   |   |   |   |--- time >  109.00
|   |   |   |   |   |--- weights: [5.00, 4.00] class: 0
|   |   |   |--- cumulative_time >  224.50
|   |   |   |   |--- cumulative_time <= 1126.50
|   |   |   |   |   |--- weights: [51.00, 48.00] class: 0
|   |   |   |   |--- cumulative_time >  1126.50
|   |   |   |   |   |--- weights: [6.00, 0.00] class: 0
|   |   |--- time >  140.50
|   |   |   |--- time <= 172.50
|   |   |   |   |--- time <= 171.50
|   |   |   |   |   |--- weights: [15.00, 1.00] class: 0
|   |   |   |   |--- time >  171.50
|   |   |   |   |   |--- weights: [0.00, 1.00] class: 1
|   |   |   |--- time >  172.50
|   |   |   |   |--- weights: [17.00, 0.00] class: 0
```

clf_time_evals_export

```
|--- response_evals <= 0.50
|   |--- time <= 13.50
|   |   |--- weights: [319.00, 0.00] class: 0
|   |--- time >  13.50
|   |   |--- time <= 68.00
|   |   |   |--- time <= 32.00
|   |   |   |   |--- cumulative_time <= 786.50
|   |   |   |   |   |--- weights: [29.00, 0.00] class: 0
|   |   |   |   |--- cumulative_time >  786.50
|   |   |   |   |   |--- weights: [13.00, 4.00] class: 0
|   |   |   |--- time >  32.00
|   |   |   |   |--- cumulative_time <= 71.50
|   |   |   |   |   |--- weights: [8.00, 0.00] class: 0
|   |   |   |   |--- cumulative_time >  71.50
|   |   |   |   |   |--- weights: [17.00, 16.00] class: 0
|   |   |--- time >  68.00
|   |   |   |--- weights: [24.00, 0.00] class: 0
|--- response_evals >  0.50
|   |--- time <= 51.50
|   |   |--- time <= 32.50
|   |   |   |--- cumulative_time <= 490.50
|   |   |   |   |--- cumulative_evals <= 18.50
|   |   |   |   |   |--- weights: [0.00, 17.00] class: 1
|   |   |   |   |--- cumulative_evals >  18.50
|   |   |   |   |   |--- weights: [1.00, 0.00] class: 0
|   |   |   |--- cumulative_time >  490.50
|   |   |   |   |--- weights: [0.00, 38.00] class: 1
|   |   |--- time >  32.50
|   |   |   |--- cumulative_evals <= 36.00
|   |   |   |   |--- cumulative_time <= 229.50
|   |   |   |   |   |--- weights: [0.00, 5.00] class: 1
|   |   |   |   |--- cumulative_time >  229.50
|   |   |   |   |   |--- weights: [16.00, 45.00] class: 1
|   |   |   |--- cumulative_evals >  36.00
|   |   |   |   |--- weights: [2.00, 0.00] class: 0
|   |--- time >  51.50
|   |   |--- time <= 140.50
|   |   |   |--- cumulative_time <= 230.50
|   |   |   |   |--- time <= 55.50
|   |   |   |   |   |--- weights: [0.00, 2.00] class: 1
|   |   |   |   |--- time >  55.50
|   |   |   |   |   |--- weights: [23.00, 5.00] class: 0
|   |   |   |--- cumulative_time >  230.50
|   |   |   |   |--- cumulative_time <= 1126.50
|   |   |   |   |   |--- weights: [42.00, 48.00] class: 1
|   |   |   |   |--- cumulative_time >  1126.50
|   |   |   |   |   |--- weights: [4.00, 0.00] class: 0
|   |   |--- time >  140.50
|   |   |   |--- cumulative_evals <= 23.50
|   |   |   |   |--- cumulative_evals <= 2.50
|   |   |   |   |   |--- weights: [4.00, 1.00] class: 0
|   |   |   |   |--- cumulative_evals >  2.50
|   |   |   |   |   |--- weights: [16.00, 0.00] class: 0
|   |   |   |--- cumulative_evals >  23.50
|   |   |   |   |--- weights: [0.00, 1.00] class: 1
```

Figure 5.15. Classification model data.

userid	bktquizid	attemptid	firstname	lastname	email	questionsattempted	masteryacheived	questionid	mastery	timetaken
558	22	6628				5	1	12637	0.208929154	151
558	22	6628				5	1	12636	0.942288432	137
558	22	6628				5	1	12662	0.796995669	313
558	22	6628				5	1	12635	0.942544638	94
558	22	6628				5	1	12626	0.957768982	45

Figure 5.16. BKT Report example.

When a student initiates the adaptive assessment, a question is randomly selected from the question pool. For assessment using enhanced BKT models, the student's answer is classified based on time spent on the task and the number of code evaluations. By considering both the answer correctness and the classification result, the BKT API calculates the updated knowledge mastery probability, which is then reflected in the BKT-Q prototype module. The assessment concludes once the student's knowledge mastery probability reaches the predefined threshold.

The Moodle LMS prototype modules and the BKT API are available upon request.

5.4 In situ experimental guidelines

Subsection 5.4.1 provides insights into SP and adaptive formative assessments, while subsection 5.4.2 presents the results of a usability study on the student learning experience and the programming environments used in formative assessments.

5.4.1 Self-practice and adaptive formative assessment

The enhanced BKT models for cognitive modelling proposed in this thesis were developed using data from a large sample of undergraduate students. This subsection further explored the relationship between the CE and SP assessments completed a day before the laboratory exercises. Understanding this relationship provided insights that could inform future research on formative assessment strategies.

The correlation between CE and SP assessment results was examined using the Pearson r correlation measure and the corresponding p-value (Table 5.5).

Table 5.5. The relationship between SP and CE formative assessments.

CE Column	SP Column	Correlation	p-value	Mean CE Grade (%)	SD CE Grade (%)	Mean SP Grade (%)	SD SP Grade (%)	# Students
DKC_CE_01	DKC_SP_01	0.63507	0.00000	41.13636	31.98174	20.00000	24.97125	88
DKC_CE_02	DKC_SP_02	0.62022	0.00000	55.70588	30.78490	34.00000	26.56752	85
DKC_CE_03	DKC_SP_03	0.74357	0.00000	35.28169	23.17461	16.69014	17.13275	71
DKC_CE_04	DKC_SP_04	0.64470	0.00000	49.91667	35.68330	21.41667	31.55694	60
DKC_CE_05	DKC_SP_05	0.70521	0.00000	20.42373	18.50648	9.91525	13.97324	59
DKC_CE_06	DKC_SP_06	0.78791	0.00000	32.15517	35.04523	19.39655	27.14609	58
DKC_CE_07	DKC_SP_07	0.86473	0.00000	9.90566	21.93720	6.41509	17.46913	53
DKC_CE_08	DKC_SP_08	0.72436	0.00000	63.15574	30.46608	38.93443	28.14682	122
DKC_CE_09	DKC_SP_09	0.70902	0.00000	23.96694	23.66359	13.84298	18.24327	121
DKC_CE_10	DKC_SP_10	0.64575	0.00000	35.66667	25.60145	22.79167	23.92913	120
DKC_CE_11	DKC_SP_11	0.72700	0.00000	65.28302	32.62945	48.06604	34.63119	106
DKC_CE_12	DKC_SP_12	0.72100	0.00000	32.42857	34.90450	22.23810	30.93092	105
DKC_CE_13	DKC_SP_13	0.69708	0.00000	40.45455	32.35715	21.91919	25.35168	99
DKC_CE_14	DKC_SP_14	0.66804	0.00000	34.38144	31.74818	15.30928	22.60179	97
DKC_CE_15	DKC_SP_15	0.76036	0.00000	49.59770	33.02763	31.55172	30.55029	87
DKC_CE_16	DKC_SP_16	0.70000	0.00000	62.87356	38.06842	32.06897	34.98797	87
DKC_CE_17	DKC_SP_17	0.80590	0.00000	23.10345	30.54963	11.83908	22.82536	87
DKC_CE_19	DKC_SP_19	0.63159	0.00000	67.27273	33.14813	36.98864	36.78565	88
DKC_CE_21	DKC_SP_21	0.54663	0.00000	27.81250	31.97707	14.63542	23.45483	96
DKC_CE_23	DKC_SP_23	0.55242	0.00000	61.35870	39.92509	29.72826	38.61876	92
DKC_CE_25	DKC_SP_25	0.66605	0.00000	52.28571	35.38285	27.21429	32.74446	70

The results demonstrated positive correlations between CE and SP grades across all DKCs. Several DKCs exhibited moderately strong correlations (e.g., DKC_21 and DKC_23) and very strong correlations (e.g., DKC_07 and DKC_17), while other DKCs showed strong correlations.

The average grades for CE assessments were consistently higher than those for SP assessments, suggesting that the SP assessments were generally more challenging.

A potential limitation in addressing the cold start problem of BKT model fitting using SP assessments lies in the contextual differences between SP and CE environments. The practice strategies specific to the SP context may not directly apply to the CE. However, the SP data can serve as a valuable proxy for assessing students' prior knowledge within the CE framework. Future research should consider hybrid models that integrate data from both contexts to improve model accuracy and student performance predictions.

In addition, the results of the adaptive formative assessments used to test the Moodle LMS prototype modules were observed. The four CE adaptive assessments (DKC 18 Lists - Basic 22/23 CE, DKC 20 Sorting algorithm CE, DKC 22 Files - Basic 22/23 CE, and DKC 24 Files including Lists CE) were fitted using data from the earlier experimenting or the related SP assessments. The vanilla BKT model and the enhanced BKT modelling were examined. Table 5.6 presents the descriptive statistics of the results related to the number of answer opportunities each student reached in the adaptive assessment. The statistics included the mean (%), Standard Deviation (SD), median (%), minimum (%) and maximum (%) values of answer opportunities, and the number of students that achieved knowledge mastery (# Mastery achieved).

Table 5.6. Adaptive formative assessments – the central tendency and dispersion measures.

Domain Knowledge Component	Students		Answer opportunities			BKT KM probability	
	#	# KM achieved	Mean %	SD %	Median %	Mean %	SD %
DKC 18 Lists - Basic 22/23 CE	82	55 (67%)	11.61	5.31	9	0.65	0.43
DKC 20 Sorting algorithm CE G1-G2	44	36 (81%)	9.14	5.20	6.5	0.78	0.28
DKC 20 Sorting algorithm CE G3-G5	57	33 (57%)	10.58	7.74	10	0.60	0.44
DKC 22 Files - Basic 22/23 CE	99	42 (42%)	14.61	3.81	14	0.67	0.36
DKC 24 Files including Lists CE G1-G2	29	18 (62%)	12.41	3.26	11	0.81	0.26
DKC 24 Files including Lists CE G3-G5	46	15 (32%)	16.37	3.53	19	0.52	0.40

DKC 20 Sorting Algorithm CE G1-G2 resulted in the highest number of students reaching knowledge mastery at 81%, while DKC 24 Files CE G3-G5 referred to the lowest 32%. Median answer opportunities ranged from 6.5 to 19, with the highest variability in DKC 20 Sorting Algorithm CE G3-G5. The BKT probabilities align with these trends, with the highest mean of the knowledge mastery probability in DKC 24 Files including Lists CE G1-G2 (0.81) and the lowest in the G3-G5 (0.52).

From the student's perspective, the adaptive assessments primarily demonstrated the idea of time efficiency in the adaptive assessment. Based on the overall learning experience, students participated in the usability study.

5.4.2 Usability study

At the end of the semester, a usability study was conducted to assess students' experiences with formative assessments and the related environment. The study employed an anonymous questionnaire via Google Forms, consisting of various questions and responses on a 7-point Likert scale. The questionnaire gathered general student information, experiences with SP and CE assessments (Appendix F), and feedback on using the VPL-based assessment environment within Moodle LMS (Appendix G). For the latter case, the System Usability Scale (SUS) [129, 130] was applied, a widely recognized instrument for assessing software product usability [131]. The SUS score was derived from responses to a ten-item questionnaire based on a 5-point Likert scale.

A total of 91 students participated in the usability study. Regarding self-evaluation computer skills, 66% of students reported above-average digital literacy, and 89% indicated above-average experience with e-learning systems like Moodle LMS. However, 47% of students reported no prior experience with basic programming concepts, and 60% had no experience with Python.

In terms of learning materials used for self-study, 65% of students reported using weekly assessments, 64% relied on presentations from laboratory exercises, 56% watched lecture videos, 56% watched videos from laboratory exercises, 48% referred to the course book, 44% used lecture presentations, 33% utilized personal notes from laboratory exercises, 25% referred to personal lecture notes and 22% reported using other sources.

Regarding weekly self-study time, 14.3% of students reported spending more than 6 hours per week, 39.6% indicated they did not study on their own, and 46.1% reported studying less than 6 hours per week.

The percentage of over-average answers on the 5-point Likert scale (points 3 to 5) was analysed for the SP formative assessments. Overall, 90% of students reported feeling comfortable during the assessments, 85% acknowledged a positive impact on their learning performance, and 78% indicated that the assessments helped them improve their knowledge of basic programming concepts. Additionally, 90% of students recommended the continued use of these assessments in the course, while 69% suggested using shorter adaptive assessments. Furthermore, 93% of students supported using bonus points based on assessment results, and 74% favoured scheduling assessments one day before laboratory exercises.

In response to the timing of the SP assessments, 49.4% of students preferred not to limit the time for completing the tests, suggesting an entire week until the following laboratory exercises. Meanwhile, 28.6% of students chose a day before the laboratory exercises and 22% preferred two days before.

A similar analysis was conducted for the CE assessments, showing that 86% of students (a -4% decrease compared to SP assessments) reported feeling comfortable during the assessments. Additionally, 87% of students (a +2% increase) agreed on the positive impact on their learning performance, and 84% (a +6% increase) reported improved knowledge of basic programming concepts. 91% of students (a +1% increase) recommended using the assessments in the course, while 76% of students (a +7% increase) suggested using shorter adaptive assessments. Furthermore, 95% of students (a +2% increase) supported using bonus points based on CE assessment performance.

Regarding the VPL-based assessment environment within Moodle LMS, the SUS score of 77.17033 significantly exceeded the industry standard threshold of 68. This result confirmed the robust usability of the approach and highlighted its user-centric design.

6 CONCLUSION

While enhanced BKT models demonstrated strong predictive capabilities, further research was needed to fully understand their potential in estimating knowledge mastery across a broader range of educational domains. This gap was particularly evident in areas like introductory programming, where studies remained scarce. This doctoral thesis proposed a novel approach for adaptive formative assessment through enhanced BKT modelling and introduced a framework for model ranking based on student performance prediction, knowledge mastery estimation, and model convergence.

The study successfully demonstrated the feasibility of using BKT models to track student knowledge by incorporating the features of time spent on task and the number of code evaluations. These features proved significant for student modelling in the introductory programming course and were effectively integrated into BKT-based cognitive models.

Student performance prediction was improved by using the enhanced BKT models, which included time and evaluation data. These models outperformed the baseline vanilla BKT model in predicting student performance across multiple Domain Knowledge Components (DKCs). The enhanced models achieved lower RMSE and higher AUC scores, both of which were statistically significant, indicating more accurate predictions of student performance.

The study also showed that enhanced BKT models outperformed the vanilla BKT model in estimating students' knowledge mastery. This was evident from higher Pearson correlation coefficients and F1 scores, showing that the enhanced models provided a more precise assessment of how well students had mastered the concept.

The enhanced BKT models, particularly those incorporating features such as time spent on task and the number of code evaluations, provided more effective and reliable paths to knowledge mastery than the baseline BKT model. Faster model convergence indicated a reduced number of answer opportunities were required for students to achieve knowledge mastery. This efficiency underscored the practical advantages of enhanced BKT models in adaptive learning systems, their applicability to real-time formative assessments, and their ability to enhance student engagement and educational outcomes.

The research results supported the development of a framework for ranking BKT models based on their capacity to predict student performance, estimate knowledge mastery, and model efficient learning paths.

This thesis made the following scientific contributions:

- A novel approach for adaptive formative assessment in the programming domain, enhancing the baseline Bayesian Knowledge Tracing model by incorporating time spent on task and the number of code evaluations

  The integration of the proposed features provided a more fine-grained understanding of a student's progress and engagement. The enhanced BKT models improved the performance results of the vanilla BKT model.

- A framework for ranking BKT model variants based on their ability to predict student performance and estimate knowledge mastery effectively

  This framework offered a systematic approach to identifying and selecting the most effective BKT models, ensuring more accurate assessments compared to the vanilla BKT model.

An additional result of this research, though not considered a scientific contribution, is:

- A publicly available dataset developed for formative assessment within the Introductory Programming course, encompassing features related to the time spent on task and the number of code evaluations.

Future research should address certain limitations of in situ experimental settings and explore extensions to the proposed approach. Specifically, the accuracy of the time on task feature requires improvement and additional features, such as the number of code runs, should be investigated. Moreover, incorporating new data sources, such as self-practice assessments, could further enhance model accuracy and robustness. Regarding the proposed model evaluation approach, the slightly lower performance compared to cross-validation suggests potential overfitting, highlighting the need for further model tuning and a more diverse dataset to improve generalization. Additionally, alternative evaluation methods for limited datasets should be explored (e.g., bootstrapping, a resampling technique that leverages the original dataset).

In conclusion, both the BKT model ranking framework and usability study results indicated that the enhanced BKT models provided a more refined understanding of student knowledge mastery. These models showed improved cognitive modelling capabilities and represented more effective choices for predicting student performance and mastery. Additionally, integrating these advanced BKT models into LMS can enhance their adaptive and predictive functionalities and provide more personalized learning experiences. The publicly available dataset for formative assessment in the programming domain, encompassing features such as time spent on task and the number of code evaluations, represents a valuable contribution to the field, supporting further advancements in educational data science.

BIBLIOGRAPHY

[1] I. Šarić Grgić, A. Grubišić and A. Gašpar, Twenty-Five Years of Bayesian knowledge tracing: a systematic review, *User Modeling and User-Adapted Interaction*, 2024.

[2] K. Zhang and Y. Yao, A three learning states Bayesian knowledge tracing model, *Knowledge Based Systems*, 148, 189–201, 2018.

[3] M. Yudelson, K. Koedinger and G. Gordon, Individualized Bayesian Knowledge Tracing Models, H. Lane, K. Yacef, J. Mostow and P. Pavlik, editors, *Artificial Intelligence in Education*, Lecture Notes in Computer Science, 171–180, Springer Berlin Heidelberg, 2013.

[4] Overview of the Moodle question engine - MoodleDocs https://docs.moodle.org/dev/Overview_of_the_moodle_question_engine Accessed on 2024-11-04.

[5] Moodle Plugins directory: Virtual Programming Lab https://moodle.org/plugins/mod_vpl Accessed on 2024-11-03, Jun. 2024.

[6] Moodle Plugins directory: VPL Question https://moodle.org/plugins/qtype_vplquestion Accessed on 2024-11-04, Apr. 2024.

[7] R. Baker, Data Mining for Education., *International Encyclopedia of Education (3rd edition). McGaw, B., Peterson, P., Baker, E. (Eds.)*, Oxford, UK: Elsevier., 2010.

[8] R. Baker and P. Inventado, Educational Data Mining and Learning Analytics, J. Larusson and B. White, editors, *Learning Analytics: From Research to Practice*, 61–75, Springer, New York, NY, 2014.

[9] A. Corbett and J. R. Anderson, Knowledge tracing: Modeling the acquisition of procedural knowledge, *User Modeling and User-Adapted Interaction*, 4, 4, 253–278, 1995.

[10] N. Hoić-Božić and M. Holenko Dlab, *Uvod u e-učenje: obrazovni izazovi digitalnog doba*, Sveučilište u Rijeci, Odjel za informatiku, 2021.

[11] B. F. Skinner, The science of learning and the art of teaching, *Harvard Educational Review*, 24, 86–97, 1954, place: US Publisher: Harvard Education Publishing Group.

[12] D. Sleeman and J. S. Brown, Introduction: Intelligent Tutoring Systems: An Overview, *Intelligent Tutoring Systems*, 1–11, Academic Press, Burlington, MA, sleeman, d.h., brown, j.s. edn., 1982.

[13] E. Wenger, *Artificial Intelligence and Tutoring Systems*, Morgan Kaufmann Publishers, Inc., California, USA, 1987.

[14] J. A. DeFalco and A. M. Sinatra, Adaptive Instructional Systems: The Evolution of Hybrid Cognitive Tools and Tutoring Systems, R. A. Sottilare and J. Schwarz, editors, *Adaptive Instructional Systems*, 11597, 52–61, Springer International Publishing, Cham, 2019, series Title: Lecture Notes in Computer Science.

[15] M. Anouar Tadlaoui, A. Souhaib, M. Khaldi and R. Carvalho, Learner Modeling in Adaptive Educational Systems: A Comparative Study, *International Journal of Modern Education and Computer Science*, 8, 1–10, 2016.

[16] P. Brusilovsky, Adaptive navigation support, *The adaptive web*, 263–290, Springer-Verlag, 2007.

[17] K. Chrysafiadi and M. Virvou, Student modeling approaches: A literature review for the last decade, *Expert Systems with Applications*, 40, 11, 4715–4729, Sep. 2013.

[18] M. C. Desmarais and R. S. J. d. Baker, A review of recent advances in learner and skill modeling in intelligent learning environments, *User Modeling and User-Adapted Interaction*, 22, 1, 9–38, Apr. 2012.

[19] B. Harrison and D. Roberts, A Review of Student Modeling Techniques in Intelligent Tutoring Systems, *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 8, 61–66, 2012, number: 5.

[20] L. Kurup, A. Joshi and N. Shekhokar, A review on student modeling approaches in ITS, *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2513–2517, 2016.

[21] Q. Liu, S. Shen, Z. Huang, E. Chen and Y. Zheng, A Survey of Knowledge Tracing, *arXiv:2105.15106 [cs]*, 2021, arXiv: 2105.15106.

[22] P. Pavlik, K. Brawner, A. Olney and A. Mitrovic, A Review of Learner Models Used in Intelligent Tutoring Systems, *Design Recommendations for Intelligent Tutoring Systems - Learner Modeling*, Volume I, 39–68, Army Research Labs, University of Memphis, 2013.

[23] R. Pelánek, Bayesian Knowledge Tracing, Logistic Models, and Beyond: An Overview of Learner Modeling Techniques, *User Modeling and User-Adapted Interaction*, 27, 3-5, 313–350, Dec. 2017.

[24] S. Ramírez Luelmo, N. El Mawas and J. Heutte, Machine Learning Techniques for Knowledge Tracing: A Systematic Literature Review:, *Proceedings of the 13th International Conference on Computer Supported Education*, 60–70, SCITEPRESS - Science and Technology Publications, Online Streaming, — Select a Country —, 2021.

[25] S. Sani, A. Bichi and S. Ayuba, Artificial intelligence approaches in student modeling: half decade review (2010-2015), *International Journal of Computer Science and Network*, 5, 5, 746–754, 2016, number: 5 Publisher: D-LAR Labs (Digital Library of Academic Research).

[26] M. Vandewaetere, P. Desmet and G. Clarebout, The contribution of learner characteristics in the development of computer-based adaptive learning environments, *Computers in Human Behavior*, 27, 1, 118–130, Jan. 2011.

[27] A. Zafar and N. Ahmad, An overview of Student Modeling Approaches under Uncertain Conditions, *International Journal of Information Technology and Management*, 4, 1, 2013.

[28] J. R. Anderson, *Rules of the mind*, Rules of the mind, Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, 1993, pages: ix, 320.

[29] R. C. Atkinson, Optimizing the learning of a second-language vocabulary, *Journal of Experimental Psychology*, 96, 1, 124–129, 1972, place: US Publisher: American Psychological Association.

[30] R. Baker, A. Corbett, S. Gowda, A. Wagner, B. MacLaren, L. Kauffman, A. Mitchell and S. Giguere, Contextual Slip and Prediction of Student Performance after Use of an Intelligent Tutor, P. Bra, A. Kobsa and D. Chin, editors, *User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings*, 6075 of *Lecture Notes in Computer Science*, 52–63, Springer, Big Island, HI, USA, 2010.

[31] J. E. Beck and J. Sison, Using Knowledge Tracing to Measure Student Reading Proficiencies, J. C. Lester, R. M. Vicari and F. Paraguaçu, editors, *Intelligent Tutoring Systems*, 624–634, Springer, Berlin, Heidelberg, 2004.

[32] C. Lin, S. Shen and M. Chi, Incorporating Student Response Time and Tutor Instructional Interventions into Student Modeling, J. Vassileva, J. Blustein, L. Aroyo and S. D'Mello, editors, *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, UMAP 2016, Halifax, NS, Canada, July 13 - 17, 2016*, 157–161, ACM, 2016.

[33] C. Lin and M. Chi, Intervention-BKT: Incorporating Instructional Interventions into Bayesian Knowledge Tracing, A. Micarelli, J. Stamper and K. Panourgia, editors, *Intelligent Tutoring Systems*, Lecture Notes in Computer Science, 208–218, Springer International Publishing, Cham, 2016.

[34] T. Schodde, K. Bergmann and S. Kopp, Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making, B. Mutlu, M. Tscheligi, A. Weiss and J. Young, editors, *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI 2017, Vienna, Austria, March 6-9, 2017*, 128–136, ACM, Vienna, Austria, 2017.

[35] M. Yudelson, O. Medvedeva and R. Crowley, A multifactor approach to student model evaluation, *User Modeling and User-Adapted Interaction*, 18, 4, 349–382, 2008.

[36] J. González-Brenes, Y. Huang and P. Brusilovsky, General Features in Knowledge Tracing to Model Multiple Subskills, Temporal Item Response Theory, and Expert Knowledge, J. Stamper, Z. Pardos, M. Mavrikis and B. McLaren, editors, *Proceedings of the 7th International Conference on Educational Data Mining, EDM 2014, London, UK, July 4-7, 2014*, 84–91, International Educational Data Mining Society (IEDMS), London, UK, 2014.

[37] M. Khajah, Y. Huang, J. P. González-Brenes, M. C. Mozer and P. Brusilovsky, Integrating knowledge tracing and item response theory: A tale of two frameworks,

CEUR Workshop Proceedings, 1181, 7–15, University of Pittsburgh, Jan. 2014, iSSN: 1613-0073.

[38] Y. Huang, J. Guerra and P. Brusilovsky, Modeling Skill Combination Patterns for Deeper Knowledge Tracing, 2016.

[39] Y. Huang and P. Brusilovsky, Towards Modeling Chunks in a Knowledge Tracing Framework for Students' Deep Learning, T. Barnes, M. Chi and M. Feng, editors, *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, 666–668, International Educational Data Mining Society (IEDMS), 2016.

[40] S. Wang, Y. Han, W. Wu and Z. Hu, Modeling student learning outcomes in studying programming language course, *2017 Seventh International Conference on Information Science and Technology (ICIST)*, 263–270, 2017.

[41] W. F. Atchison, S. D. Conte, J. W. Hamblen, T. E. Hull, T. A. Keenan, W. B. Kehl, E. J. McCluskey, S. O. Navarro, W. C. Rheinboldt, E. J. Schweppe, W. Viavant and D. M. Young, Curriculum 68: Recommendations for academic programs in computer science: a report of the ACM curriculum committee on computer science, *Communications of the ACM*, 11, 3, 151–197, Mar. 1968.

[42] P. Brusilovsky, B. J. Ericson, C. S. Horstmann, C. Servin, F. Vahid and C. Zilles, The Future of Computing Education Materials, First Draft, to be published in the CS2023: ACM/IEEE-CS/AAAI Computer Science Curricula., Technical Report. ACM, 2023.

[43] M. H. Brown and R. Sedgewick, Techniques for Algorithm Animation, *IEEE Software*, 2, 1, 28–39, Jan. 1985.

[44] S. Benford, E. Burke and E. Foxley, Learning to construct quality software with the Ceilidh system, *Software Quality Journal*, 2, 3, 177–197, Sep. 1993.

[45] J. R. Anderson and B. J. Reiser, LISP TUTOR., *Byte*, 10, 4, Apr. 1985.

[46] R. Pelánek, Adaptive, Intelligent, and Personalized: Navigating the Terminological Maze Behind Educational Technology, *International Journal of Artificial Intelligence in Education*, 32, 1, 151–173, Mar. 2022.

[47] K. Vanlehn, The Behavior of Tutoring Systems, *International Journal of Artificial Intelligence in Education*, 16, 3, 227–265, Aug. 2006.

[48] A. Essa, A possible future for next generation adaptive learning systems, *Smart Learning Environments*, 3, 1, 16, Nov. 2016.

[49] V. Aleven, E. McLaughlin, R. Glenn and K. Koedinger, Instruction Based on Adaptive Learning Technologies, *In R. E. Mayer & P. Alexander (Eds.), Handbook of research on learning and instruction*, 522–560, Routledge, 2 edn., 2016, num Pages: 39.

[50] M. Liu, V. Rus and L. Liu, Automatic Chinese Factual Question Generation, *IEEE Transactions on Learning Technologies*, 10, 2, 194–204, Apr. 2017, conference Name: IEEE Transactions on Learning Technologies.

[51] R. L. Bangert-Drowns, J. A. Kulik and C.-L. C. Kulik, Effects of Frequent Classroom Testing, *The Journal of Educational Research*, 85, 2, 89–99, 1991, publisher: Taylor & Francis, Ltd.

[52] D. H. Smith, C. Emeka, M. Fowler, M. West and C. Zilles, Investigating the Effects of Testing Frequency on Programming Performance and Students' Behavior, *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2023, 757–763, Association for Computing Machinery, New York, NY, USA, Mar. 2023.

[53] D. Garcia, C. McMahon, Y. Garcia, M. West and C. Zilles, Achieving "A's for All (as Time and Interest Allow)", *Proceedings of the Ninth ACM Conference on Learning @ Scale*, L@S '22, 255–258, Association for Computing Machinery, New York, NY, USA, Jun. 2022.

[54] C. Kulik, J. Kulik and R. L. Bangert-Drowns, Effectiveness of Mastery Learning Programs: A Meta-Analysis, *Review of Educational Research*, 60, 2, 265–299, Jun. 1990, publisher: American Educational Research Association.

[55] J. W. Morphew, M. Silva, G. Herman and M. West, Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering, *Applied Cognitive Psychology*, 34, 1, 168–181, 2020, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.3605.

[56] J. Williamson, Cognitive Diagnostic Models and How They Can Be Useful. Research Report, Tech. rep., Cambridge University Press & Assessment, Dec. 2023, eRIC Number: ED639603.

[57] T. Mitchell, *Machine Learning*, McGraw-Hill Education, New York, 1997.

[58] H. Cen, K. Koedinger and B. Junker, Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement, M. Ikeda, K. D. Ashley and T.-W. Chan, editors, *Intelligent Tutoring Systems*, Lecture Notes in Computer Science, 164–175, Springer, Berlin, Heidelberg, 2006.

[59] P. I. Pavlik, H. Cen and K. R. Koedinger, Performance Factors Analysis –A New Alternative to Knowledge Tracing, *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, 531–538, IOS Press, NLD, Jul. 2009.

[60] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas and J. Sohl-Dickstein, Deep Knowledge Tracing, *Advances in Neural Information Processing Systems*, 28, Curran Associates, Inc., 2015.

[61] S. Montero, A. Arora, S. Kelly, B. Milne and M. Mozer, Does deep knowledge tracing model interactions among skills?, 2018.

[62] A. Badrinath, F. Wang and Z. Pardos, pyBKT: An Accessible Python Library of Bayesian Knowledge Tracing Models, *CoRR*, abs/2105.00385, 2021, arXiv: 2105.00385.

[63] M. Khajah, R. Lindsey and M. Mozer, How Deep is Knowledge Tracing?, T. Barnes, M. Chi and M. Feng, editors, *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, International Educational Data Mining Society (IEDMS), Raleigh, North Carolina, USA, 2016.

[64] D. Moher, A. Liberati, J. Tetzlaff, D. Altman and T. P. Group, Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement, *PLOS Medicine*, 6, 7, e1000097, 2009, publisher: Public Library of Science.

[65] K. Chang, J. Beck, J. Mostow and A. Corbett, Does Help Help? A Bayes Net Approach to Modeling Tutor Interventions, Boston, Massachusetts, 2006.

[66] R. Baker, A. Corbett and V. Aleven, Improving Contextual Models of Guessing and Slipping with a Truncated Training Set, Carnegie Mellon University, 2008.

[67] J. E. Beck, K.-m. Chang, J. Mostow and A. Corbett, Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology, B. P. Woolf, E. Aïmeur, R. Nkambou and S. Lajoie, editors, *Intelligent Tutoring Systems*, 383–394, Springer, Berlin, Heidelberg, 2008.

[68] R. Baker, A. Corbett and V. Aleven, More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing, B. Woolf, E. Aïmeur, R. Nkambou and S. Lajoie, editors, *Proceedings of the 9th Inernational Conference on Intelligent Tutoring Systems, ITS 2008, Montreal, Canada, June 23-27, 2008*, 5091 of *Lecture Notes in Computer Science*, 406–415, Springer, Montreal, Canada, 2008.

[69] D. Halpern, S. Tubridy, H. Wang, C. Gasser, P. Popp, L. Davachi and T. Gureckis, Knowledge Tracing Using the Brain, K. Boyer and M. Yudelson, editors, *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*, International Educational Data Mining Society (IEDMS), Buffalo, NY, USA, 2018.

[70] F. Liu, X. Hu, C. Bu and K. Yu, Fuzzy Bayesian Knowledge Tracing, *IEEE Transactions on Fuzzy Systems*, 1–1, 2021, conference Name: IEEE Transactions on Fuzzy Systems.

[71] Y. David, A. Segal and Y. Gal, Sequencing educational content in classrooms using Bayesian knowledge tracing, D. Gasevic, G. Lynch, S. Dawson, H. Drachsler and C. Rosé, editors, *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, LAK 2016, Edinburgh, United Kingdom, April 25-29, 2016*, 354–363, ACM, Edinburgh, United Kingdom, 2016.

[72] K. Ostrow, C. Donnelly, S. Adjei and N. Heffernan, Improving Student Modeling Through Partial Credit and Problem Difficulty, G. Kiczales, D. Russell and B. Woolf, editors, *Proceedings of the Second ACM Conference on Learning @ Scale, L@S 2015, Vancouver, BC, Canada, March 14 - 18, 2015*, 11–20, ACM, Vancouver, BC, Canada, 2015.

[73] Y. Wang, N. Heffernan and J. Beck, Representing Student Performance with Partial Credit, R. Baker, A. Merceron and P. Pavlik, editors, *Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13, 2010. Proceedings*, 335–336, www.educationaldatamining.org, Pittsburgh, PA, USA, 2010.

[74] Y. Wang and N. Heffernan, Extending Knowledge Tracing to Allow Partial Credit: Using Continuous versus Binary Nodes, H. Lane, K. Yacef, J. Mostow and P. Pavlik, editors, *Artificial Intelligence in Education - 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings*, 7926 of *Lecture Notes in Computer Science*, 181–188, Springer, Memphis, TN, USA, 2013.

[75] Z. Wang, J. Zhu, X. Li, Z. Hu and M. Zhang, Structured Knowledge Tracing Models for Student Assessment on Coursera, J. Haywood, V. Aleven, J. Kay and I. Roll, editors, *Proceedings of the Third ACM Conference on Learning @ Scale, L@S 2016, Edinburgh, Scotland, UK, April 25 - 26, 2016*, 209–212, ACM, Edinburgh, Scotland, UK, 2016.

[76] M. Eagle, A. Corbett, J. Stamper, B. McLaren, A. Wagner, B. MacLaren and A. Mitchell, Estimating Individual Differences for Student Modeling in Intelligent Tutors from Reading and Pretest Data, A. Micarelli, J. Stamper and K. Panourgia, editors, *Intelligent Tutoring Systems - 13th International Conference, ITS 2016, Zagreb, Croatia, June 7-10, 2016. Proceedings*, 9684 of *Lecture Notes in Computer Science*, 133–143, Springer, Zagreb, Croatia, 2016.

[77] M. Eagle, A. Corbett, J. Stamper, B. McLaren, R. Baker, A. Wagner, B. MacLaren and A. Mitchell, Predicting Individual Differences for Learner Modeling in Intelligent Tutors from Previous Learner Activities, J. Vassileva, J. Blustein, L. Aroyo and S. D'Mello, editors, *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, UMAP 2016, Halifax, NS, Canada, July 13 - 17, 2016*, 55–63, ACM, Halifax, NS, Canada, 2016.

[78] M. Eagle, A. Corbett, J. Stamper, B. McLaren, R. Baker, A. Wagner, B. MacLaren and A. Mitchell, Exploring Learner Model Differences Between Students, E. André, R. Baker, X. Hu, M. Rodrigo and B. Du Boulay, editors, *Artificial Intelligence in Education - 18th International Conference, AIED 2017, Wuhan, China, June 28 - July 1, 2017, Proceedings*, 10331 of *Lecture Notes in Computer Science*, 494–497, Springer, Wuhan, China, 2017.

[79] P. Nedungadi and M. Remya, Predicting students' performance on intelligent tutoring system - Personalized clustered BKT (PC-BKT) model, *IEEE Frontiers in Education Conference, FIE 2014, Proceedings, Madrid, Spain, October 22-25, 2014*, 1–6, IEEE Computer Society, Madrid, Spain, 2014.

[80] P. Nedungadi and M. Remya, Incorporating forgetting in the Personalized, Clustered, Bayesian Knowledge Tracing (PC-BKT) model, *2015 International Conference on Cognitive Computing and Information Processing(CCIP)*, 1–5, 2015.

[81] Z. Pardos and N. Heffernan, Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing, P. De Bra, A. Kobsa and D. Chin, editors, *User Modeling, Adaptation, and Personalization*, Lecture Notes in Computer Science, 255–266, Springer Berlin Heidelberg, 2010.

[82] Y. Song, Y. Jin, X. Zheng, H. Han, Y. Zhong and X. Zhao, PSFK: A Student Performance Prediction Scheme for First-Encounter Knowledge in ITS, S. Zhang, M. Wirsing and Z. Zhang, editors, *Knowledge Science, Engineering and Management - 8th International Conference, KSEM 2015, Chongqing, China, October 28-30, 2015, Proceedings*, 9403 of *Lecture Notes in Computer Science*, 639–650, Springer, Chongqing, China, 2015.

[83] L. Wang, J. Jiang, H. L. Chieu, C. H. Ong, D. Song and L. Liao, Can Syntax Help? Improving an LSTM-based Sentence Compression Model for New Domains, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1385–1393, Association for Computational Linguistics, Vancouver, Canada, Jul. 2017.

[84] Y. Xu and J. Mostow, Using Item Response Theory to Refine Knowledge Tracing, S. K. D'Mello, R. A. Calvo and A. Olney, editors, *Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee, USA, July 6-9, 2013*, 356–357, International Educational Data Mining Society, 2013.

[85] D. Agarwal, N. Babel and R. Baker, Contextual Derivation of Stable BKT Parameters for Analysing Content Efficacy, K. Boyer and M. Yudelson, editors, *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*, International Educational Data Mining Society (IEDMS), Buffalo, NY, USA, 2018.

[86] Z. Pardos and N. Heffernan, KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model, J. Konstan, R. Conejo, J. Marzo and N. Oliver, editors, *User Modeling, Adaption and Personalization - 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings*, 6787 of *Lecture Notes in Computer Science*, 243–254, Springer, Girona, Spain, 2011.

[87] X. Zhou, W. Wu and Y. Han, Modeling multiple subskills by extending knowledge tracing model using logistic regression, *2017 IEEE International Conference on Big Data (Big Data)*, 3994–4003, Dec. 2017.

[88] Y. Qiu, Y. Qi, H. Lu, Z. Pardos and N. Heffernan, Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing, *EDM*, 2011.

[89] S. Adjei, S. Salehizadeh, Y. Wang and N. T. Heffernan, Do students really learn an equal amount independent of whether they get an item correct or wrong?, S. K. D'Mello, R. A. Calvo and A. Olney, editors, *Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee, USA, July 6-9, 2013*, 304–305, International Educational Data Mining Society, 2013.

[90] R. Baker, S. Gowda and E. Salamin, Modeling the Learning That Takes Place Between Online Assessments, *Proceedings of the 26th International Conference on Computers in Education, Philippines: Asia-Pacific Society for Computers in EducationAsia-Pacific Society for Computers in Education*, 8, Asia-Pacific Society for Computers in Education, Philippines, 2018.

[91] M. Sao Pedro, R. Baker and J. Gobert, Incorporating Scaffolding and Tutor Context into Bayesian Knowledge Tracing to Predict Inquiry Skill Acquisition, S. D'Mello,

R. Calvo and A. Olney, editors, *Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee, USA, July 6-9, 2013*, 185–192, International Educational Data Mining Society, Memphis, Tennessee, USA, 2013.

[92] K. I. Chan, R. Tse and P. I. Lei, Tracing Students' Learning Performance on Multiple Skills using Bayesian Methods, *Proceedings of the 6th International Conference on Education and Multimedia Technology*, ICEMT '22, 84–89, Association for Computing Machinery, New York, NY, USA, Nov. 2022.

[93] W. Hawkins, N. Heffernan and R. Baker, Learning Bayesian Knowledge Tracing Parameters with a Knowledge Heuristic and Empirical Probabilities, *12th International Conference on Intelligent Tutoring Systems - Volume 8474*, ITS 2014, 150–155, Springer-Verlag, Berlin, Heidelberg, Jun. 2014.

[94] Y. Huang and L. He, Automatic generation of short answer questions for reading comprehension assessment, *Natural Language Engineering*, 22, 3, 457–489, May 2016, publisher: Cambridge University Press.

[95] Z. MacHardy, Applications of bayesian knowledge tracing to the curation of educational videos, Tech. rep., University of California at Berkeley, 2015.

[96] Z. MacHardy and Z. Pardos, Toward the Evaluation of Educational Videos using Bayesian Knowledge Tracing and Big Data, G. Kiczales, D. Russell and B. Woolf, editors, *Proceedings of the Second ACM Conference on Learning @ Scale, L@S 2015, Vancouver, BC, Canada, March 14 - 18, 2015*, 347–350, ACM, 2015.

[97] L. Meng, M. Zhang, W. Zhang and Y. Chu, CS-BKT: introducing item relationship to the Bayesian knowledge tracing model, *Interactive Learning Environments*, 1–11, Jun. 2019.

[98] S. Sun, X. Hu, C. Bu, F. Liu, Y. Zhang and W. Luo, Genetic Algorithm for Bayesian Knowledge Tracing: A Practical Application, Y. Tan, Y. Shi and B. Niu, editors, *Advances in Swarm Intelligence*, Lecture Notes in Computer Science, 282–293, Springer International Publishing, Cham, 2022.

[99] M. Khajah, R. Wing, R. Lindsey and M. Mozer, Incorporating Latent Factors Into Knowledge Tracing To Predict Individual Differences In Learning, 2014.

[100] Z. Pardos, Y. Bergner, D. Seaton and D. Pritchard, Adapting Bayesian Knowledge Tracing to a Massive Open Online Course in edX, S. D'Mello, R. Calvo and A. Olney, editors, *Proceedings of the 6th International Conference on Educational Data Mining, Memphis, Tennessee, USA, July 6-9, 2013*, 137–144, International Educational Data Mining Society, Memphis, Tennessee, USA, 2013.

[101] S. Bhatt, J. Zhao, C. Thille, D. Zimmaro and N. Gattani, Evaluating Bayesian Knowledge Tracing for Estimating Learner Proficiency and Guiding Learner Behavior, D. Joyner, R. Kizilcec and S. Singer, editors, *L@S'20: Seventh ACM Conference on Learning @ Scale, Virtual Event, USA, August 12-14, 2020*, 357–360, ACM, Virtual , USA, 2020.

[102] S. Corrigan, T. Barkley and Z. Pardos, Dynamic Approaches to Modeling Student Affect and its Changing Role in Learning and Performance, F. Ricci, K. Bontcheva, O. Conlan and S. Lawless, editors, *User Modeling, Adaptation and Personalization - 23rd International Conference, UMAP 2015, Dublin, Ireland, June 29 - July 3, 2015. Proceedings*, 9146 of *Lecture Notes in Computer Science*, 92–103, Springer, Dublin, Ireland, 2015.

[103] M. Eagle, A. Corbett, J. Stamper and B. McLaren, Predicting Individualized Learner Models Across Tutor Lessons, K. Boyer and M. Yudelson, editors, *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*, International Educational Data Mining Society (IEDMS), Buffalo, NY, USA, 2018.

[104] G. Gorgun and O. Bulut, Considering Disengaged Responses in Bayesian and Deep Knowledge Tracing, M. M. Rodrigo, N. Matsuda, A. I. Cristea and V. Dimitrova, editors, *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium*, Lecture Notes in Computer Science, 591–594, Springer International Publishing, Cham, 2022.

[105] J. Lee and E. Brunskill, The Impact on Individualizing Student Models on Necessary Practice Opportunities, K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson and J. Stamper, editors, *Proceedings of the 5th International Conference on Educational Data Mining, Chania, Greece, June 19-21, 2012*, 118–125, www.educationaldatamining.org, Chania, Greece, 2012.

[106] Z. Pardos, S. Trivedi, N. Heffernan and G. Sárközy, Clustered Knowledge Tracing, S. Cerri, W. Clancey, G. Papadourakis and K. Panourgia, editors, *Intelligent Tutoring Systems - 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18, 2012. Proceedings*, 7315 of *Lecture Notes in Computer Science*, 405–410, Springer, Chania, Crete, Greece, 2012.

[107] M. Rau and Z. Pardos, Adding eye-tracking AOI data to models of representation skills does not improve prediction accuracy, T. Barnes, M. Chi and M. Feng, editors, *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016*, 622–623, International Educational Data Mining Society (IEDMS), 2016.

[108] S. Spaulding, G. Gordon and C. Breazeal, Affect-Aware Student Models for Robot Tutors, C. Jonker, S. Marsella, J. Thangarajah and K. Tuyls, editors, *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, May 9-13, 2016*, 864–872, ACM, 2016.

[109] Y. Wang and N. Heffernan, The Student Skill Model, S. Cerri, W. Clancey, G. Papadourakis and K. Panourgia, editors, *Intelligent Tutoring Systems - 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18, 2012. Proceedings*, 7315 of *Lecture Notes in Computer Science*, 399–404, Springer, Chania, Crete, Greece, 2012.

[110] M. Falakmasir, M. Yudelson, S. Ritter and K. Koedinger, Spectral Bayesian Knowledge Tracing, O. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron,

P. Mitros, J. Luna, M. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura and M. Desmarais, editors, *Proceedings of the 8th International Conference on Educational Data Mining, EDM 2015, Madrid, Spain, June 26-29, 2015*, 360–363, International Educational Data Mining Society (IEDMS), Madrid, Spain, 2015.

[111] J. E. Beck, Difficulties in inferring student knowledge from observations (and why you should care, 2007.

[112] J. Beck and K. Chang, Identifiability: A Fundamental Problem of Student Modeling, C. Conati, K. McCoy and G. Paliouras, editors, *User Modeling 2007, 11th International Conference, UM 2007, Corfu, Greece, June 25-29, 2007, Proceedings*, 4511 of *Lecture Notes in Computer Science*, 137–146, Springer, Corfu, Greece, 2007.

[113] Moodle Home https://moodle.org/ Accessed on 2024-11-03.

[114] VPL - Virtual Programming Lab - Home https://vpl.dis.ulpgc.es/ Accessed on 2024-11-03.

[115] J. C. Rodríguez-del Pino, Z. J. Hernández-Figueroa, J. D. G. Domínguez and J. Skalka, Virtual Programming Lab for Moodle: Automatic Program Assessment in a First-year University Course, E. Smyrnova-Trybulska, P. Kommers, M. Drlík and J. Skalka, editors, *Microlearning: New Approaches To A More Effective Higher Education*, 195–206, Springer International Publishing, Cham, 2022.

[116] D. Thiébaut, Automatic evaluation of computer programs using Moodle's virtual programming lab (VPL) plug-in, *Journal of Computing Sciences in Colleges*, 30, 6, 145–151, Jun. 2015.

[117] A. Von Wangenheim, J. Martina, R. Cancian and J. Dovicchi, *Developing Programming Courses with Moodle and VPL - The Teacher's Guide to the Virtual Programming Lab*, Dec. 2015.

[118] S. Doroudi and E. Brunskill, The Misidentified Identifiability Problem of Bayesian Knowledge Tracing, Tech. rep., International Educational Data Mining Society, Jun. 2017, publication Title: International Educational Data Mining Society ERIC Number: ED596611.

[119] Quiz report statistics - MoodleDocs https://docs.moodle.org/dev/Quiz_report_statistics Accessed on 2024-11-03.

[120] Kolmogorov-Smirnov Calculator (Test of Normality) https://www.socscistatistics.com/tests/kolmogorov/ Accessed on 2024-11-03.

[121] I. Šarić Grgić, File repository https://shorturl.at/Y7V67 Accessed on 2024-11-03.

[122] Wilcoxon Signed-Rank Test https://www.socscistatistics.com/tests/signedranks/ Accessed on 2024-11-06.

[123] Moodle (v4.01) Database https://www.examulator.com 2024-11-06/er/4.1/ Accessed on 2024-11-06.

[124] Moodle Plugins directory: Adaptive Quiz: CAT (Computer-Adaptive Testing) implementation for Moodle https://moodle.org/plugins/mod_adaptivequiz Accessed on 2024-11-04, Apr. 2024.

[125] B. Wright, Practical Adaptive Testing CAT Algorithm, *Rasch Measurement Transactions*, Rasch Measurement Transactions, p.24.

[126] J. Linacre, Computer-Adaptive Testing: A Methodology Whose Time Has Come, 2000.

[127] Adaptive Quiz: middlebury/moodle-mod_adaptivequiz https://github.com/middlebury/moodle-mod_adaptivequiz Accessed on 2024-11-06.

[128] Moodle Plugins directory: ALiSe CAT Quiz https://moodle.org/plugins/local_catquiz Accessed on 2024-11-06, Oct. 2024.

[129] J. Brooke, SUS: A 'Quick and Dirty' Usability Scale, *Usability Evaluation In Industry*, CRC Press, 1996, num Pages: 6.

[130] J. Brooke, SUS: A Retrospective - JUX, *JUX - The Journal of User Experience*, 8, 2, pp. 29–40, 2013.

[131] A. Bangor, P. T. Kortum and J. T. Miller, An Empirical Evaluation of the System Usability Scale, *International Journal of Human–Computer Interaction*, 24, 6, 574–594, Jul. 2008, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447310802205776.

APPENDIX A

BKT model parameter probabilities – DKC 01-25

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.36265	0.12490		0.29319		0.00905					
#02 vanilla+forgets	0.41225	0.09530		0.26573		0.01742			0.03109		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.56526	0.03282	0.13562	0.50000	0.24735	0.00932					
#04 multigs+forgets T	0.41959	0.04399	0.20317	0.33860	0.12116	0.03861			0.05923		
#05 multilearn T	0.38431	0.12453		0.29911		0.00445	0.01438				
#06 multilearn+forgets T	0.39864	0.11094		0.25062		0.01014	0.02353		0.02926	0.02569	
#07 multigs+multilearn T	0.57115	0.03272	0.13557	0.50000	0.25256	0.00557	0.02044				
#08 multigs+multilearn+forgets T	0.46941	0.02807	0.14253	0.37749	0.14819	0.04718	0.07464		0.06418	0.07013	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.13838		0.26961		0.13520	0.42771	0.02356			
#10 multiprior+forgets T		0.12396		0.19579		0.13645	0.46764	0.03553	0.00000	0.00000	0.04805
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.55251	0.03255	0.10478	0.50000	0.30691	0.01098					
#12 multigs+forgets TE	0.63112	0.02304	0.07863	0.50000	0.30452	0.01170			0.01732		
#13 multilearn TE	0.38351	0.12458		0.29883		0.00465	0.01076				
#14 multilearn+forgets TE	0.40923	0.10582		0.25267		0.01150	0.02114		0.03714	0.02720	
#15 multigs+multilearn TE	0.57768	0.03193	0.09918	0.50000	0.30912	0.00492	0.01492				
#16 multigs+multilearn+forgets TE	0.50653	0.02141	0.08027	0.47395	0.18500	0.07268	0.11531		0.11114	0.07891	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.13838		0.26961		0.07944	0.37367	0.02356			
#18 multiprior+forgets TE		0.12396		0.19579		0.07927	0.39661	0.03553	0.00000	0.00000	0.04805

Table A1. BKT model parameter probabilities – DKC 01.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.48397	0.20352		0.24283		0.00897					
#02 vanilla+forgets	0.53543	0.16325		0.19964		0.02237			0.02641		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.73350	0.02613	0.08045	0.50000	0.36721	0.02914					
#04 multigs+forgets T	0.80619	0.00374	0.00590	0.50000	0.36715	0.05698			0.02186		
#05 multilearn T	0.46846	0.22188		0.22972		0.00280	0.00747				
#06 multilearn+forgets T	0.55069	0.18209		0.18514		0.00869	0.02126		0.05419	0.02803	
#07 multigs+multilearn T	0.73519	0.02559	0.07647	0.50000	0.36704	0.02674	0.03337				
#08 multigs+multilearn+forgets T	0.82205	0.00475	0.01738	0.50000	0.35007	0.04432	0.08825		0.06496	0.02644	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.25944		0.18654		0.05685	0.21909	0.03175			
#10 multiprior+forgets T		0.17335		0.16040		0.13210	0.32025	0.06419	0.00000	0.00000	0.04873
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.71887	0.03193	0.10294	0.50000	0.33378	0.02705					
#12 multigs+forgets TE	0.80129	0.00312	0.00646	0.50000	0.32816	0.06037			0.02277		
#13 multilearn TE	0.47485	0.21041		0.23726		0.00415	0.01137				
#14 multilearn+forgets TE	0.55361	0.17899		0.18868		0.01013	0.02189		0.05029	0.02528	
#15 multigs+multilearn TE	0.72258	0.02744	0.08106	0.50000	0.33372	0.03512	0.03381				
#16 multigs+multilearn+forgets TE	0.80264	0.00746	0.03177	0.50000	0.30857	0.04889	0.07442		0.04896	0.02285	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.25944		0.18654		0.05685	0.21909	0.03175			
#18 multiprior+forgets TE		0.17335		0.16040		0.13210	0.32025	0.06419	0.00000	0.00000	0.04873

Table A2. BKT model parameter probabilities – DKC 02.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.58225	0.08448		0.50000		0.00328					
#02 vanilla+forgets	0.62074	0.05443		0.39953		0.03189			0.06279		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.65411	0.03449	0.09435	0.50000	0.44436	0.00316					
#04 multigs+forgets T	0.74227	0.00000	0.00000	0.50000	0.43407	0.05604			0.05979		
#05 multilearn T	0.56592	0.09268		0.50000		0.00258	0.00286				
#06 multilearn+forgets T	0.64190	0.06078		0.40113		0.02055	0.02919		0.06507	0.05579	
#07 multigs+multilearn T	0.65023	0.03454	0.09426	0.50000	0.44426	0.00392	0.00408				
#08 multigs+multilearn+forgets T	0.77167	0.00169	0.00857	0.50000	0.40452	0.03591	0.06131		0.06844	0.05055	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.19292		0.37418		0.09330	0.17254	0.02146			
#10 multiprior+forgets T		0.07256		0.29487		0.18773	0.51011	0.08671	0.00000	0.00000	0.12528
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.62835	0.04430	0.43987	0.50000	0.12147	0.00443					
#12 multigs+forgets TE	0.64820	0.02359	0.31008	0.46384	0.09621	0.04423			0.06502		
#13 multilearn TE	0.58057	0.08814		0.50000		0.00207	0.00164				
#14 multilearn+forgets TE	0.63073	0.05098		0.40083		0.03460	0.02353		0.06751	0.04649	
#15 multigs+multilearn TE	0.64142	0.04122	0.42574	0.50000	0.14088	0.00380	0.00328				
#16 multigs+multilearn+forgets TE	0.65695	0.01219	0.19989	0.46838	0.09719	0.06436	0.04747		0.07775	0.06261	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.19292		0.37418		0.05635	0.23142	0.02146			
#18 multiprior+forgets TE		0.07256		0.29487		0.14459	0.51837	0.08671	0.00000	0.00000	0.12528

Table A3. BKT model parameter probabilities – DKC 03.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.39733	0.05934		0.24948		0.01703					
#02 vanilla+forgets	0.40679	0.04880		0.22847		0.02234			0.01469		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.52069	0.01342	0.04326	0.50000	0.25949	0.01678					
#04 multigs+forgets T	0.39508	0.01251	0.05117	0.33386	0.14776	0.05786			0.05878		
#05 multilearn T	0.40269	0.06258		0.24780		0.01186	0.02342				
#06 multilearn+forgets T	0.40841	0.05537		0.22115		0.01529	0.03061		0.01913	0.01420	
#07 multigs+multilearn T	0.51776	0.01128	0.03490	0.50000	0.26319	0.01546	0.02795				
#08 multigs+multilearn+forgets T	0.57315	0.00836	0.01092	0.50000	0.23973	0.01459	0.02636		0.01774	0.00855	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.06889		0.23115		0.24772	0.31122	0.02924			
#10 multiprior+forgets T		0.06212		0.18191		0.25404	0.31296	0.03810	0.00000	0.00000	0.03061
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.52468	0.01126	0.04792	0.50000	0.25946	0.01701					
#12 multigs+forgets TE	0.56892	0.00452	0.01759	0.50000	0.25695	0.01761			0.00916		
#13 multilearn TE	0.40288	0.05940		0.25049		0.01466	0.02101				
#14 multilearn+forgets TE	0.40186	0.05106		0.21069		0.02341	0.03215		0.03681	0.02183	
#15 multigs+multilearn TE	0.52361	0.00911	0.03602	0.50000	0.26184	0.01837	0.02133				
#16 multigs+multilearn+forgets TE	0.57288	0.00352	0.01444	0.50000	0.23697	0.02065	0.02596		0.02681	0.01079	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.06889		0.23115		0.09126	0.76521	0.02924			
#18 multiprior+forgets TE		0.06212		0.18191		0.09292	0.77862	0.03810	0.00000	0.00000	0.03061

Table A4. BKT model parameter probabilities – DKC 04.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.29136	0.08448		0.50000		0.00655					
#02 vanilla+forgets	0.46387	0.03798		0.50000		0.02917			0.07085		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.33665	0.04728	0.50000	0.50000	0.14536	0.01092					
#04 multigs+forgets T	0.50326	0.01705	0.50000	0.50000	0.17096	0.04666			0.09034		
#05 multilearn T	0.29331	0.08481		0.50000		0.00607	0.02154				
#06 multilearn+forgets T	0.45365	0.04131		0.50000		0.02655	0.06898		0.06905	0.03197	
#07 multigs+multilearn T	0.30671	0.04721	0.50000	0.50000	0.13990	0.01308	0.04116				
#08 multigs+multilearn+forgets T	0.50344	0.01674	0.50000	0.50000	0.16318	0.04726	0.11420		0.09214	0.08389	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.08845		0.50000		0.08351	1.00000	0.02536			
#10 multiprior+forgets T		0.08305		0.48790		0.09816	1.00000	0.03635	0.00000	0.00000	0.07155
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.31690	0.06558	0.30583	0.50000	0.34616	0.00882					
#12 multigs+forgets TE	0.50128	0.02051	0.11280	0.50000	0.30018	0.04258			0.08569		
#13 multilearn TE	0.29621	0.08535		0.50000		0.00548	0.01371				
#14 multilearn+forgets TE	0.45857	0.04190		0.50000		0.02553	0.07051		0.07339	0.05651	
#15 multigs+multilearn TE	0.29910	0.06660	0.32308	0.50000	0.33943	0.00948	0.02589				
#16 multigs+multilearn+forgets TE	0.49370	0.03004	0.16316	0.50000	0.31405	0.02774	0.07621		0.07615	0.05730	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.08845		0.50000		0.08351	1.00000	0.02536			
#18 multiprior+forgets TE		0.08305		0.48790		0.09816	1.00000	0.03635	0.00000	0.00000	0.07155

Table A5. BKT model parameter probabilities – DKC 05.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.44240	0.01019		0.39336		0.00793					
#02 vanilla+forgets	0.43373	0.00335		0.34843		0.01559			0.01479		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.47079	0.00455	0.03692	0.50000	0.29441	0.00606					
#04 multigs+forgets T	0.43933	0.00056	0.00435	0.45142	0.25651	0.02109			0.02139		
#05 multilearn T	0.45362	0.01126		0.39452		0.00527	0.01273				
#06 multilearn+forgets T	0.42435	0.00790		0.34953		0.01051	0.02169		0.00563	0.01371	
#07 multigs+multilearn T	0.40483	0.00756	0.05714	0.49453	0.27847	0.00997	0.02200				
#08 multigs+multilearn+forgets T	0.50942	0.00000	0.00004	0.50000	0.27454	0.00797	0.02319		0.01118	0.01279	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.01468		0.37206		0.30279	1.00000	0.02234			
#10 multiprior+forgets T		0.01015		0.30098		0.30581	1.00000	0.03242	0.00000	0.00000	0.02883
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.41158	0.01143	0.01983	0.47322	0.22395	0.01132					
#12 multigs+forgets TE	0.44098	0.00193	0.00850	0.42850	0.21594	0.01722			0.01701		
#13 multilearn TE	0.43676	0.01066		0.39143		0.00898	0.00085				
#14 multilearn+forgets TE	0.43156	0.00393		0.34800		0.01596	0.00671		0.01452	0.01411	
#15 multigs+multilearn TE	0.43563	0.00800	0.01425	0.48510	0.22510	0.01096	0.00180				
#16 multigs+multilearn+forgets TE	0.44381	0.00143	0.00664	0.42920	0.21576	0.01816	0.01075		0.01857	0.01591	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.01468		0.37206		0.30279	1.00000	0.02234			
#18 multiprior+forgets TE		0.01015		0.30098		0.30581	1.00000	0.03242	0.00000	0.00000	0.02883

Table A6. BKT model parameter probabilities – DKC 06.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.13101	0.01164		0.45353		0.00020					
#02 vanilla+forgets	0.14437	0.01033		0.42631		0.00058			0.01663		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.13579	0.00922	0.50000	0.48386	0.00000	0.00044					
#04 multigs+forgets T	0.15644	0.00678	0.50000	0.46110	0.00000	0.00099			0.01968		
#05 multilearn T	0.13431	0.01111		0.46045		0.00013	0.00001				
#06 multilearn+forgets T	0.13624	0.01128		0.41862		0.00055	0.00004		0.00938	0.18241	
#07 multigs+multilearn T	0.13575	0.00917	0.50000	0.48456	0.00000	0.00048	0.00007				
#08 multigs+multilearn+forgets T	0.15541	0.00685	0.50000	0.46384	0.00000	0.00086	0.00010		0.01420	0.16994	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.01242		0.41949		0.06161	1.00000	0.00500			
#10 multiprior+forgets T		0.01357		0.35505		0.05752	1.00000	0.00562	0.00000	0.00000	0.02005
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.13438	0.00906	0.19863	0.49350	0.11110	0.00054					
#12 multigs+forgets TE	0.16066	0.00664	0.19896	0.48456	0.11166	0.00040			0.01640		
#13 multilearn TE	0.13428	0.01111		0.46043		0.00013	0.00000				
#14 multilearn+forgets TE	0.13681	0.01122		0.41712		0.00063	0.00003		0.01077	0.06982	
#15 multigs+multilearn TE	0.13786	0.00892	0.19926	0.49682	0.11115	0.00026	0.00001				
#16 multigs+multilearn+forgets TE	0.15886	0.00667	0.19772	0.48138	0.11157	0.00067	0.00007		0.01445	0.06008	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.01242		0.41949		0.06161	1.00000	0.00500			
#18 multiprior+forgets TE		0.01357		0.35505		0.05752	1.00000	0.00562	0.00000	0.00000	0.02005

Table A7. BKT model parameter probabilities – DKC 07.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.45216	0.29955		0.11806		0.02018					
#02 vanilla+forgets	0.49184	0.26031		0.10554		0.04135			0.02133		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.83736	0.08021	0.14633	0.50000	0.20447	0.01033					
#04 multigs+forgets T	0.96975	0.01051	0.02465	0.50000	0.21149	0.02165			0.01110		
#05 multilearn T	0.42913	0.30726		0.11107		0.02182	0.02099				
#06 multilearn+forgets T	0.51366	0.25798		0.10606		0.04093	0.04121		0.02831	0.02257	
#07 multigs+multilearn T	0.84225	0.08134	0.14889	0.50000	0.20487	0.00441	0.01129				
#08 multigs+multilearn+forgets T	0.96110	0.02135	0.04904	0.50000	0.20863	0.01233	0.02799		0.01932	0.00946	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.33327		0.09444		0.13096	0.26440	0.04436			
#10 multiprior+forgets T		0.29982		0.07823		0.15873	0.29755	0.06514	0.00000	0.00000	0.02636
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.84072	0.07250	0.16266	0.50000	0.19155	0.01084					
#12 multigs+forgets TE	0.62824	0.09459	0.28350	0.25700	0.05310	0.15113			0.09740		
#13 multilearn TE	0.43530	0.30855		0.11124		0.01683	0.02318				
#14 multilearn+forgets TE	0.50750	0.27051		0.10695		0.02807	0.03834		0.02299	0.01748	
#15 multigs+multilearn TE	0.84297	0.07311	0.16397	0.50000	0.19161	0.00680	0.01399				
#16 multigs+multilearn+forgets TE	0.95692	0.01833	0.04726	0.50000	0.19108	0.02142	0.04163		0.02172	0.00985	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.33327		0.09444		0.13637	0.27282	0.04436			
#18 multiprior+forgets TE		0.29982		0.07823		0.16196	0.30838	0.06514	0.00000	0.00000	0.02636

Table A8. BKT model parameter probabilities – DKC 08.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.32417	0.08216		0.50000		0.00019					
#02 vanilla+forgets	0.42328	0.07405		0.34000		0.00288			0.08115		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.35845	0.04938	0.19462	0.50000	0.38614	0.00004					
#04 multigs+forgets T	0.59606	0.00746	0.05275	0.50000	0.38098	0.05926			0.14415		
#05 multilearn T	0.32167	0.08302		0.50000		0.00012	0.00013				
#06 multilearn+forgets T	0.42969	0.07433		0.33374		0.00353	0.00391		0.09902	0.06555	
#07 multigs+multilearn T	0.35806	0.04938	0.19458	0.50000	0.38597	0.00011	0.00011				
#08 multigs+multilearn+forgets T	0.59539	0.01908	0.12930	0.50000	0.27246	0.01543	0.01790		0.09719	0.07583	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.14041		0.33266		0.04407	0.22649	0.00532			
#10 multiprior+forgets T		0.08372		0.25075		0.13011	0.44151	0.02453	0.00000	0.00000	0.12269
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.34485	0.06214	0.25954	0.50000	0.29204	0.00012					
#12 multigs+forgets TE	0.54976	0.04229	0.19362	0.50000	0.13649	0.00198			0.07109		
#13 multilearn TE	0.32178	0.08302		0.50000		0.00010	0.00016				
#14 multilearn+forgets TE	0.44164	0.07502		0.33338		0.00248	0.00669		0.11400	0.05798	
#15 multigs+multilearn TE	0.34453	0.06225	0.25985	0.50000	0.29210	0.00009	0.00011				
#16 multigs+multilearn+forgets TE	0.48709	0.02634	0.15156	0.48249	0.10233	0.06489	0.13284		0.30250	0.15605	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.14041		0.33266		0.03691	0.21485	0.00532			
#18 multiprior+forgets TE		0.08372		0.25075		0.10906	0.45458	0.02453	0.00000	0.00000	0.12269

Table A9. BKT model parameter probabilities – DKC 09.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.65433	0.05922		0.50000		0.00012					
#02 vanilla+forgets	0.76249	0.02693		0.41416		0.01216			0.05608		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.66825	0.04213	0.12645	0.50000	0.35151	0.00027					
#04 multigs+forgets T	0.78157	0.00537	0.03683	0.48685	0.28995	0.02621			0.06039		
#05 multilearn T	0.65197	0.06071		0.50000		0.00005	0.00002				
#06 multilearn+forgets T	0.74252	0.03205		0.41545		0.00770	0.01394		0.03750	0.08864	
#07 multigs+multilearn T	0.66910	0.04165	0.12552	0.50000	0.36410	0.00035	0.00013				
#08 multigs+multilearn+forgets T	0.76980	0.00480	0.03270	0.48624	0.29139	0.02575	0.04794		0.04899	0.09138	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.21631		0.37343		0.12062	1.00000	0.01556			
#10 multiprior+forgets T		0.06328		0.30787		0.37990	1.00000	0.07368	0.00000	0.00000	0.11480
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.66916	0.04564	0.26361	0.50000	0.10305	0.00027					
#12 multigs+forgets TE	0.81069	0.00433	0.05096	0.50000	0.10151	0.01468			0.05007		
#13 multilearn TE	0.65188	0.06082		0.50000		0.00002	0.00004				
#14 multilearn+forgets TE	0.76894	0.02675		0.41175		0.01326	0.01916		0.06609	0.04715	
#15 multigs+multilearn TE	0.66952	0.04559	0.26256	0.50000	0.10304	0.00022	0.00058				
#16 multigs+multilearn+forgets TE	0.65138	0.01717	0.33951	0.46921	0.04565	0.13927	0.24939		0.26863	0.19515	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.21631		0.37343		0.05845	0.23930	0.01556			
#18 multiprior+forgets TE		0.06328		0.30787		0.22685	0.67209	0.07368	0.00000	0.00000	0.11480

Table A10. BKT model parameter probabilities – DKC 10.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.59551	0.15206		0.17735		0.03333					
#02 vanilla+forgets	0.53672	0.17374		0.12554		0.05054			0.01770		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.84200	0.00000	0.00074	0.50000	0.19931	0.03032					
#04 multigs+forgets T	0.84392	0.00000	0.00000	0.50000	0.19529	0.03032			0.00498		
#05 multilearn T	0.57774	0.16942		0.16964		0.02238	0.04011				
#06 multilearn+forgets T	0.54270	0.18529		0.11916		0.04015	0.05515		0.02743	0.01893	
#07 multigs+multilearn T	0.77968	0.00334	0.03232	0.50000	0.20624	0.03636	0.03911				
#08 multigs+multilearn+forgets T	0.86136	0.00000	0.00000	0.50000	0.16938	0.02116	0.03616		0.01482	0.00562	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.20401		0.15283		0.20117	0.43483	0.06641			
#10 multiprior+forgets T		0.20656		0.09245		0.19510	0.42771	0.08814	0.00000	0.00000	0.03385
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.81171	0.01624	0.03290	0.50000	0.16553	0.02355					
#12 multigs+forgets TE	0.82434	0.00930	0.01782	0.50000	0.16546	0.02686			0.00319		
#13 multilearn TE	0.57513	0.16887		0.16955		0.02586	0.04130				
#14 multilearn+forgets TE	0.54368	0.18954		0.12446		0.03439	0.05044		0.01751	0.01476	
#15 multigs+multilearn TE	0.80636	0.01619	0.03236	0.50000	0.16528	0.02398	0.02841				
#16 multigs+multilearn+forgets TE	0.83985	0.00447	0.00904	0.50000	0.15589	0.03075	0.03559		0.01342	0.00522	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.20401		0.15283		0.14524	0.56620	0.06641			
#18 multiprior+forgets TE		0.20656		0.09245		0.14112	0.55646	0.08814	0.00000	0.00000	0.03385

Table A11. BKT model parameter probabilities – DKC 11.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.36937	0.06581		0.33963		0.00034					
#02 vanilla+forgets	0.49917	0.01673		0.28667		0.01108			0.04984		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.46430	0.02254	0.06030	0.50000	0.21885	0.00182					
#04 multigs+forgets T	0.59660	0.00340	0.00861	0.50000	0.18985	0.00462			0.03907		
#05 multilearn T	0.36834	0.06664		0.33817		0.00015	0.00026				
#06 multilearn+forgets T	0.49312	0.01804		0.28729		0.01124	0.00653		0.04098	0.05264	
#07 multigs+multilearn T	0.47710	0.02053	0.05524	0.50000	0.21882	0.00143	0.00037				
#08 multigs+multilearn+forgets T	0.59617	0.00349	0.00902	0.50000	0.18636	0.00562	0.00275		0.04217	0.03777	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.07819		0.31650		0.22264	0.27709	0.01017			
#10 multiprior+forgets T		0.02940		0.24721		0.31878	0.34525	0.02276	0.00000	0.00000	0.05186
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.48824	0.01405	0.19395	0.50000	0.17068	0.00111					
#12 multigs+forgets TE	0.59776	0.00271	0.03139	0.50000	0.14939	0.00454			0.03899		
#13 multilearn TE	0.36829	0.06660		0.33818		0.00021	0.00006				
#14 multilearn+forgets TE	0.50082	0.01671		0.28824		0.01069	0.00968		0.04849	0.04974	
#15 multigs+multilearn TE	0.48919	0.01369	0.19016	0.50000	0.17080	0.00141	0.00038				
#16 multigs+multilearn+forgets TE	0.59907	0.00229	0.02427	0.50000	0.14204	0.00604	0.00228		0.04384	0.03765	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.07819		0.31650		0.18247	0.54519	0.01017			
#18 multiprior+forgets TE		0.02940		0.24721		0.25590	0.68656	0.02276	0.00000	0.00000	0.05186

Table A12. BKT model parameter probabilities – DKC 12.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.33843	0.16119		0.27053		0.00056					
#02 vanilla+forgets	0.52617	0.09411		0.20435		0.01860			0.07269		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.62026	0.03218	0.11644	0.50000	0.28440	0.00147					
#04 multigs+forgets T	0.77104	0.00221	0.01004	0.50000	0.26487	0.02275			0.04594		
#05 multilearn T	0.33210	0.16455		0.26434		0.00034	0.00053				
#06 multilearn+forgets T	0.50865	0.10303		0.19952		0.01377	0.01702		0.06314	0.07031	
#07 multigs+multilearn T	0.61374	0.03303	0.11744	0.50000	0.28413	0.00254	0.00197				
#08 multigs+multilearn+forgets T	0.76796	0.00537	0.02225	0.50000	0.26355	0.01521	0.01536		0.04466	0.03923	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.21958		0.13498		0.10330	0.14491	0.00889			
#10 multiprior+forgets T		0.10786		0.13036		0.23825	0.30477	0.04920	0.00000	0.00000	0.10684
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.63623	0.02423	0.33676	0.50000	0.18887	0.00213					
#12 multigs+forgets TE	0.77728	0.00143	0.02624	0.50000	0.18256	0.02108			0.04463		
#13 multilearn TE	0.33753	0.16360		0.26719		0.00003	0.00016				
#14 multilearn+forgets TE	0.52303	0.10015		0.18600		0.01672	0.03825		0.10036	0.06893	
#15 multigs+multilearn TE	0.60928	0.02768	0.35241	0.50000	0.18895	0.00456	0.00505				
#16 multigs+multilearn+forgets TE	0.78003	0.00355	0.06472	0.50000	0.10911	0.01893	0.04021		0.05761	0.03763	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.21958		0.13498		0.05144	0.30841	0.00889			
#18 multiprior+forgets TE		0.10786		0.13036		0.15046	0.58084	0.04920	0.00000	0.00000	0.10684

Table A13. BKT model parameter probabilities – DKC 13.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.33361	0.09566		0.35778		0.00086					
#02 vanilla+forgets	0.43377	0.05579		0.26417		0.01317			0.06169		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.34055	0.08302	0.50000	0.37107	0.00000	0.00238					
#04 multigs+forgets T	0.43943	0.05163	0.50000	0.26689	0.00000	0.01483			0.06425		
#05 multilearn T	0.33286	0.09866		0.35420		0.00010	0.00000				
#06 multilearn+forgets T	0.43231	0.05649		0.26537		0.01239	0.00269		0.05891	0.36107	
#07 multigs+multilearn T	0.33891	0.09423	0.50000	0.36724	0.00000	0.00178	0.00004				
#08 multigs+multilearn+forgets T	0.44048	0.05243	0.50000	0.27090	0.00000	0.01362	0.00291		0.06046	0.35919	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.09629		0.35091		0.21291	1.00000	0.01321			
#10 multiprior+forgets T		0.06747		0.17080		0.24355	1.00000	0.03631	0.00000	0.00000	0.10175
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.45827	0.02466	0.21169	0.50000	0.26791	0.00619					
#12 multigs+forgets TE	0.64752	0.00079	0.01106	0.50000	0.25899	0.03456			0.06687		
#13 multilearn TE	0.33302	0.09886		0.35407		0.00002	0.00009				
#14 multilearn+forgets TE	0.44287	0.05873		0.26927		0.00784	0.01805		0.06246	0.05392	
#15 multigs+multilearn TE	0.46049	0.02465	0.21278	0.50000	0.26813	0.00498	0.01216				
#16 multigs+multilearn+forgets TE	0.64255	0.00487	0.06423	0.50000	0.24396	0.02346	0.05251		0.09065	0.04755	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.09629		0.35091		0.06698	0.49471	0.01321			
#18 multiprior+forgets TE		0.06747		0.17080		0.07617	0.56677	0.03631	0.00000	0.00000	0.10175

Table A14. BKT model parameter probabilities – DKC 14.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.48230	0.10753		0.30720		0.00882					
#02 vanilla+forgets	0.53460	0.08136		0.19217		0.02348			0.04907		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.65099	0.01830	0.08374	0.50000	0.28034	0.00792					
#04 multigs+forgets T	0.74282	0.00000	0.00002	0.50000	0.27255	0.02627			0.02942		
#05 multilearn T	0.46500	0.11942		0.29492		0.00811	0.00704				
#06 multilearn+forgets T	0.53478	0.08823		0.19270		0.01593	0.02168		0.04289	0.04492	
#07 multigs+multilearn T	0.66185	0.01756	0.07958	0.50000	0.28069	0.00736	0.00446				
#08 multigs+multilearn+forgets T	0.75497	0.00079	0.00457	0.50000	0.26481	0.02211	0.03227		0.04019	0.02793	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.12688		0.28085		0.15033	0.42486	0.02878			
#10 multiprior+forgets T		0.08517		0.15930		0.20181	0.47612	0.05242	0.00000	0.00000	0.06491
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.65034	0.01818	0.31299	0.50000	0.13344	0.00788					
#12 multigs+forgets TE	0.74222	0.00025	0.00169	0.50000	0.12634	0.02629			0.02954		
#13 multilearn TE	0.46779	0.12063		0.29482		0.00666	0.00971				
#14 multilearn+forgets TE	0.54639	0.08149		0.18689		0.02394	0.03590		0.07316	0.04186	
#15 multigs+multilearn TE	0.65453	0.01772	0.30308	0.50000	0.13334	0.00817	0.00978				
#16 multigs+multilearn+forgets TE	0.75185	0.00069	0.00621	0.50000	0.10075	0.02680	0.05256		0.04313	0.02368	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.12688		0.28085		0.14568	0.56123	0.02878			
#18 multiprior+forgets TE		0.08517		0.15930		0.18488	0.63789	0.05242	0.00000	0.00000	0.06491

Table A15. BKT model parameter probabilities – DKC 15.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.50774	0.07316		0.16394		0.01417					
#02 vanilla+forgets	0.51498	0.06193		0.12155		0.02813			0.02253		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.64306	0.00609	0.02586	0.50000	0.17583	0.01411					
#04 multigs+forgets T	0.68758	0.00144	0.00564	0.50000	0.17136	0.01908			0.00896		
#05 multilearn T	0.49836	0.09017		0.15161		0.00841	0.01604				
#06 multilearn+forgets T	0.51947	0.06832		0.10894		0.02657	0.03567		0.04657	0.02484	
#07 multigs+multilearn T	0.63726	0.00601	0.02546	0.50000	0.18019	0.01335	0.01878				
#08 multigs+multilearn+forgets T	0.70305	0.00159	0.00591	0.50000	0.16812	0.01691	0.02423		0.02602	0.00968	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.10279		0.14146		0.20982	0.43785	0.02665			
#10 multiprior+forgets T		0.04141		0.11545		0.28786	0.49897	0.05330	0.00000	0.00000	0.03411
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.62170	0.01738	0.14634	0.50000	0.07287	0.01358					
#12 multigs+forgets TE	0.67938	0.00530	0.05266	0.50000	0.07343	0.01562			0.00736		
#13 multilearn TE	0.49719	0.08954		0.15178		0.01013	0.03489				
#14 multilearn+forgets TE	0.51925	0.07513		0.11240		0.01913	0.07706		0.03892	0.01964	
#15 multigs+multilearn TE	0.62796	0.01468	0.10241	0.50000	0.07287	0.01405	0.02799				
#16 multigs+multilearn+forgets TE	0.70776	0.00225	0.03315	0.50000	0.07380	0.02125	0.04238		0.04121	0.00552	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.10279		0.14146		0.09411	0.70394	0.02665			
#18 multiprior+forgets TE		0.04141		0.11545		0.11524	0.82459	0.05330	0.00000	0.00000	0.03411

Table A16. BKT model parameter probabilities – DKC 16.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.12939	0.08858		0.16681		0.00009					
#02 vanilla+forgets	0.38973	0.04046		0.19078		0.00526			0.09973		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.23784	0.03802	0.18963	0.50000	0.27128	0.00005					
#04 multigs+forgets T	0.46558	0.00792	0.06026	0.44315	0.10015	0.01981			0.12585		
#05 multilearn T	0.12925	0.08914		0.16407		0.00000	0.00001				
#06 multilearn+forgets T	0.37248	0.04952		0.18316		0.00031	0.00059		0.09523	0.06967	
#07 multigs+multilearn T	0.23252	0.03881	0.19107	0.50000	0.27136	0.00013	0.00042				
#08 multigs+multilearn+forgets T	0.60380	0.00289	0.02214	0.50000	0.20795	0.00582	0.01277		0.10682	0.06554	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.09353		0.13556		0.08021	0.13082	0.00151			
#10 multiprior+forgets T		0.04664		0.13529		0.12693	0.31482	0.01547	0.00000	0.00000	0.09869
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.26421	0.02679	0.29582	0.50000	0.29060	0.00013					
#12 multigs+forgets TE	0.59924	0.00000	0.00006	0.50000	0.25735	0.01200			0.09525		
#13 multilearn TE	0.12937	0.08907		0.16446		0.00001	0.00001				
#14 multilearn+forgets TE	0.40067	0.04761		0.17970		0.00110	0.00148		0.14894	0.07399	
#15 multigs+multilearn TE	0.26443	0.02683	0.29604	0.50000	0.29033	0.00005	0.00010				
#16 multigs+multilearn+forgets TE	0.62083	0.00076	0.01279	0.50000	0.23771	0.01058	0.02858		0.13697	0.07465	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.09353		0.13556		0.01457	0.29385	0.00151			
#18 multiprior+forgets TE		0.04664		0.13529		0.09720	0.41504	0.01547	0.00000	0.00000	0.09869

Table A17. BKT model parameter probabilities – DKC 17.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.57396	0.25482		0.11001		0.00248					
#02 vanilla+forgets	0.70156	0.17809		0.09143		0.00908			0.02232		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.82471	0.02407	0.20694	0.50000	0.20448	0.00732					
#04 multigs+forgets T	0.91283	0.01486	0.08209	0.50000	0.20448	0.02058			0.01418		
#05 multilearn T	0.57338	0.25556		0.10964		0.00184	0.00295				
#06 multilearn+forgets T	0.71186	0.20110		0.08033		0.00347	0.00868		0.06103	0.02512	
#07 multigs+multilearn T	0.83190	0.02438	0.21368	0.50000	0.20448	0.00282	0.00591				
#08 multigs+multilearn+forgets T	0.99931	0.00108	0.00672	0.50000	0.20387	0.07741	0.14602		0.13064	0.03075	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.32002		0.07679		0.00769	0.32666	0.03003			
#10 multiprior+forgets T		0.25040		0.06446		0.09439	0.42722	0.04173	0.00000	0.00000	0.02333
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.77029	0.07806	0.26870	0.50000	0.11416	0.00425					
#12 multigs+forgets TE	0.88373	0.04376	0.14291	0.50000	0.11414	0.00691			0.01172		
#13 multilearn TE	0.57289	0.25492		0.10972		0.00212	0.00391				
#14 multilearn+forgets TE	0.72214	0.18427		0.08640		0.00629	0.01332		0.05522	0.02269	
#15 multigs+multilearn TE	0.77248	0.07813	0.27292	0.50000	0.11416	0.00253	0.00532				
#16 multigs+multilearn+forgets TE	0.95733	0.03273	0.12663	0.50000	0.10979	0.02581	0.04728		0.07027	0.01302	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.32002		0.07679		0.03311	0.40378	0.03003			
#18 multiprior+forgets TE		0.25040		0.06446		0.08620	0.52434	0.04173	0.00000	0.00000	0.02333

Table A18. BKT model parameter probabilities – DKC 19.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.27822	0.08027		0.33195		0.00031					
#02 vanilla+forgets	0.45419	0.03075		0.19336		0.00658			0.07417		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.46234	0.01516	0.07300	0.50000	0.23996	0.00183					
#04 multigs+forgets T	0.46765	0.00713	0.09638	0.31964	0.09165	0.01881			0.09474		
#05 multilearn T	0.27563	0.08209		0.32742		0.00000	0.00003				
#06 multilearn+forgets T	0.45274	0.03335		0.19604		0.00283	0.01562		0.06345	0.07699	
#07 multigs+multilearn T	0.46439	0.01471	0.06926	0.50000	0.23982	0.00133	0.00771				
#08 multigs+multilearn+forgets T	0.56890	0.00409	0.04666	0.46871	0.17998	0.00192	0.01078		0.05385	0.05613	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.10054		0.26995		0.15391	0.35505	0.00591			
#10 multiprior+forgets T		0.02481		0.16537		0.30941	0.70026	0.02355	0.00000	0.00000	0.09514
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.46320	0.01568	0.14279	0.50000	0.12883	0.00185					
#12 multigs+forgets TE	0.57490	0.00558	0.05604	0.50000	0.11896	0.00305			0.05440		
#13 multilearn TE	0.27549	0.08209		0.32734		0.00002	0.00011				
#14 multilearn+forgets TE	0.45624	0.03212		0.18882		0.00500	0.02433		0.09233	0.06355	
#15 multigs+multilearn TE	0.46413	0.01534	0.13631	0.50000	0.12884	0.00196	0.00270				
#16 multigs+multilearn+forgets TE	0.59321	0.00572	0.06449	0.50000	0.11502	0.00355	0.01095		0.08601	0.03591	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.10054		0.26995		0.01978	0.50963	0.00591			
#18 multiprior+forgets TE		0.02481		0.16537		0.11140	0.85087	0.02355	0.00000	0.00000	0.09514

Table A19. BKT model parameter probabilities – DKC 21.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.60055	0.06575		0.13558		0.00096					
#02 vanilla+forgets	0.63075	0.05951		0.08833		0.00325			0.01389		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.67087	0.00989	0.02429	0.50000	0.10321	0.00019					
#04 multigs+forgets T	0.72675	0.00000	0.00000	0.50000	0.09804	0.00026			0.00817		
#05 multilearn T	0.60354	0.06657		0.13576		0.00016	0.00069				
#06 multilearn+forgets T	0.64327	0.05989		0.08317		0.00331	0.01045		0.05577	0.01337	
#07 multigs+multilearn T	0.67095	0.00989	0.02429	0.50000	0.10322	0.00013	0.00031				
#08 multigs+multilearn+forgets T	0.75560	0.00000	0.00000	0.50000	0.09835	0.00002	0.00005		0.04609	0.00698	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.11242		0.10285		0.24692	0.53676	0.01871			
#10 multiprior+forgets T		0.12120		0.03298		0.24424	0.52735	0.03223	0.00000	0.00000	0.02534
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.67082	0.01005	0.06620	0.50000	0.08818	0.00019					
#12 multigs+forgets TE	0.72667	0.00002	0.00002	0.50000	0.08536	0.00027			0.00817		
#13 multilearn TE	0.60619	0.06598		0.13662		0.00000	0.00003				
#14 multilearn+forgets TE	0.64701	0.06061		0.08443		0.00202	0.01631		0.06022	0.01206	
#15 multigs+multilearn TE	0.67097	0.01005	0.06619	0.50000	0.08818	0.00015	0.00032				
#16 multigs+multilearn+forgets TE	0.75921	0.00000	0.00001	0.50000	0.08337	0.00001	0.00004		0.05195	0.00493	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.11242		0.10285		0.03245	0.66216	0.01871			
#18 multiprior+forgets TE		0.12120		0.03298		0.05421	0.63792	0.03223	0.00000	0.00000	0.02534

Table A20. BKT model parameter probabilities – DKC 23.

BKT model	BKT parameters										
	Prior	Guess		Slip		Learn			Forgets		
#01 vanilla	0.53430	0.07072		0.24429		0.01136					
#02 vanilla+forgets	0.54052	0.06475		0.20899		0.01813			0.01579		
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#03 multigs T	0.62180	0.03384	0.13173	0.50000	0.17235	0.00547					
#04 multigs+forgets T	0.65966	0.03111	0.09522	0.50000	0.17091	0.00589			0.00677		
#05 multilearn T	0.54326	0.07153		0.24531		0.00793	0.01666				
#06 multilearn+forgets T	0.55287	0.06619		0.21190		0.01172	0.02494		0.01833	0.01349	
#07 multigs+multilearn T	0.62594	0.03409	0.13272	0.50000	0.17230	0.00308	0.01055				
#08 multigs+multilearn+forgets T	0.68893	0.03096	0.09129	0.50000	0.17705	0.00266	0.00961		0.01837	0.00734	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#09 multiprior T		0.09504		0.22438		0.24842	0.48622	0.03377			
#10 multiprior+forgets T		0.09327		0.17033		0.24825	0.50332	0.04269	0.00000	0.00000	0.02478
	Prior	Guess_Class0	Guess_Class1	Slip_Class0	Slip_Class1	Learn_Class0	Learn_Class1		Forgets_Class0	Forgets_Class1	
#11 multigs TE	0.61339	0.03228	0.28661	0.50000	0.12798	0.00720					
#12 multigs+forgets TE	0.65972	0.03064	0.19356	0.50000	0.12881	0.00574			0.00666		
#13 multilearn TE	0.53583	0.07085		0.24446		0.01131	0.00934				
#14 multilearn+forgets TE	0.54592	0.06605		0.20852		0.01553	0.03121		0.01862	0.01518	
#15 multigs+multilearn TE	0.62173	0.03214	0.29028	0.50000	0.12804	0.00604	0.00281				
#16 multigs+multilearn+forgets TE	0.68362	0.03019	0.18083	0.50000	0.12265	0.00560	0.00332		0.01824	0.00691	
	Prior	Guess		Slip		Learn_Class0	Learn_Class1	Learn_Default	Forgets_Class0	Forgets_Class1	Forgets_Default
#17 multiprior TE		0.09504		0.22438		0.14618	0.61629	0.03377			
#18 multiprior+forgets TE		0.09327		0.17033		0.14665	0.63119	0.04269	0.00000	0.00000	0.02478

Table A21. BKT model parameter probabilities – DKC 25.

APPENDIX B

Student performance prediction – DKC 01-DKC 25

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.41583	0.80246	0.74923	0.42352	0.80859	0.75667
#02 vanilla+forgets	0.40881	0.81246	0.76154	0.42125	0.81739	0.76833
#03 multigs T	0.40832	0.81508	0.73308	0.43204	0.80924	0.69667
#04 multigs+forgets T	0.40786	0.82715	0.73692	0.41446	0.82922	0.74333
#05 multilearn T	0.41546	0.80360	0.75308	0.42385	0.80843	0.75667
#06 multilearn+forgets T	0.40836	0.81396	0.75846	0.41953	0.81551	0.76333
#07 multigs+multilearn T	0.40806	0.81578	0.73385	0.43198	0.81050	0.69833
#08 multigs+multilearn+forgets T	0.40345	0.83262	0.74462	0.41935	0.82290	0.75000
#09 multiprior T	0.42643	0.78720	0.73117	0.44084	0.77283	0.72105
#10 multiprior+forgets T	0.41593	0.79921	0.75385	0.43384	0.78356	0.74211
#11 multigs TE	0.40444	0.82743	0.76538	0.42892	0.82696	0.73833
#12 multigs+forgets TE	0.40124	0.83633	0.76308	0.42401	0.83338	0.72833
#13 multilearn TE	0.41548	0.80483	0.75308	0.42382	0.80906	0.75667
#14 multilearn+forgets TE	0.40864	0.81471	0.76308	0.42036	0.81804	0.75833
#15 multigs+multilearn TE	0.40428	0.82696	0.76231	0.42893	0.82821	0.73833
#16 multigs+multilearn+forgets TE	0.40060	0.83645	0.76000	0.41856	0.82708	0.72500
#17 multiprior TE	0.42642	0.78778	0.73117	0.44025	0.77395	0.72105
#18 multiprior+forgets TE	0.41595	0.79881	0.75385	0.43315	0.78462	0.74211

Table B1. Student performance prediction – DKC 01.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.43965	0.74415	0.72500	0.42346	0.80291	0.74375
#02 vanilla+forgets	0.43047	0.76816	0.73071	0.41328	0.82858	0.75938
#03 multigs T	0.44662	0.77650	0.72643	0.43229	0.81895	0.77812
#04 multigs+forgets T	0.44173	0.78790	0.72286	0.43220	0.84290	0.77500
#05 multilearn T	0.43959	0.74379	0.72714	0.42299	0.80350	0.74688
#06 multilearn+forgets T	0.43216	0.77004	0.73286	0.41459	0.83241	0.76406
#07 multigs+multilearn T	0.44658	0.77578	0.72500	0.43217	0.81884	0.77812
#08 multigs+multilearn+forgets T	0.43679	0.78195	0.72571	0.42837	0.83516	0.75781
#09 multiprior T	0.45503	0.72895	0.69023	0.44032	0.79345	0.68586
#10 multiprior+forgets T	0.44439	0.74750	0.71579	0.42922	0.80346	0.72039
#11 multigs TE	0.44354	0.77388	0.72357	0.42248	0.84471	0.80469
#12 multigs+forgets TE	0.43915	0.77991	0.71143	0.42208	0.85319	0.77500
#13 multilearn TE	0.43967	0.74266	0.72500	0.42314	0.80373	0.74688
#14 multilearn+forgets TE	0.43173	0.76963	0.73500	0.41436	0.83264	0.76094
#15 multigs+multilearn TE	0.44490	0.77338	0.72143	0.42335	0.84369	0.80156
#16 multigs+multilearn+forgets TE	0.43392	0.77349	0.71357	0.41645	0.84887	0.77812
#17 multiprior TE	0.45503	0.72895	0.69023	0.44032	0.79345	0.68586
#18 multiprior+forgets TE	0.44439	0.74750	0.71579	0.42922	0.80346	0.72039

Table B2. Student performance prediction – DKC 02.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.43939	0.69843	0.67417	0.46192	0.62911	0.68148
#02 vanilla+forgets	0.42848	0.75087	0.71167	0.43664	0.71779	0.66111
#03 multigs T	0.43655	0.76087	0.75167	0.46390	0.70699	0.74444
#04 multigs+forgets T	0.42296	0.77861	0.73917	0.42845	0.76242	0.72778
#05 multilearn T	0.43914	0.69759	0.67417	0.46095	0.62938	0.68148
#06 multilearn+forgets T	0.42839	0.75327	0.71333	0.43614	0.71854	0.65926
#07 multigs+multilearn T	0.43655	0.76097	0.75167	0.46397	0.70693	0.74444
#08 multigs+multilearn+forgets T	0.41906	0.77967	0.73917	0.42442	0.76496	0.72778
#09 multiprior T	0.44942	0.68488	0.68947	0.45262	0.65762	0.68421
#10 multiprior+forgets T	0.43173	0.72634	0.71140	0.43914	0.69929	0.68421
#11 multigs TE	0.42739	0.73403	0.71750	0.44715	0.71854	0.75370
#12 multigs+forgets TE	0.41340	0.78470	0.70750	0.41025	0.79011	0.73704
#13 multilearn TE	0.43924	0.69761	0.67417	0.46141	0.62928	0.68148
#14 multilearn+forgets TE	0.42831	0.75310	0.71583	0.43616	0.72114	0.66296
#15 multigs+multilearn TE	0.42790	0.73409	0.71750	0.44815	0.71828	0.75370
#16 multigs+multilearn+forgets TE	0.41503	0.78242	0.71250	0.41200	0.78868	0.74630
#17 multiprior TE	0.44920	0.68539	0.69035	0.45313	0.65079	0.68421
#18 multiprior+forgets TE	0.43300	0.72417	0.71228	0.44140	0.69585	0.68421

Table B3. Student performance prediction – DKC 03.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.37710	0.84376	0.80800	0.40384	0.84169	0.77708
#02 vanilla+forgets	0.37314	0.86067	0.81500	0.39645	0.85400	0.78750
#03 multigs T	0.37506	0.85977	0.79400	0.39515	0.86730	0.78958
#04 multigs+forgets T	0.37968	0.86906	0.77600	0.38067	0.88419	0.78750
#05 multilearn T	0.37737	0.84282	0.80900	0.40291	0.84287	0.78125
#06 multilearn+forgets T	0.37324	0.85713	0.81700	0.39582	0.84669	0.78750
#07 multigs+multilearn T	0.37533	0.86095	0.79700	0.39582	0.86704	0.79375
#08 multigs+multilearn+forgets T	0.37024	0.87105	0.79100	0.38372	0.87956	0.76667
#09 multiprior T	0.38933	0.82712	0.78737	0.41257	0.83696	0.76316
#10 multiprior+forgets T	0.38563	0.84368	0.79579	0.40752	0.84451	0.75658
#11 multigs TE	0.36661	0.87088	0.80700	0.39257	0.86442	0.80000
#12 multigs+forgets TE	0.36430	0.88305	0.81500	0.38387	0.88160	0.80417
#13 multilearn TE	0.37727	0.84121	0.81000	0.40371	0.84235	0.77917
#14 multilearn+forgets TE	0.37413	0.85679	0.81500	0.39609	0.85398	0.78125
#15 multigs+multilearn TE	0.36747	0.87173	0.80500	0.39283	0.86343	0.80208
#16 multigs+multilearn+forgets TE	0.36234	0.87985	0.81300	0.37978	0.87932	0.79583
#17 multiprior TE	0.38295	0.83160	0.79789	0.40984	0.83760	0.77412
#18 multiprior+forgets TE	0.37888	0.84789	0.80632	0.40493	0.84492	0.76754

Table B4. Student performance prediction – DKC 04.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.38261	0.70629	0.80200	0.37578	0.67754	0.83182
#02 vanilla+forgets	0.37666	0.71632	0.80200	0.35310	0.74322	0.83182
#03 multigs T	0.38127	0.72323	0.80900	0.37797	0.68223	0.83409
#04 multigs+forgets T	0.37326	0.73104	0.80900	0.35240	0.74843	0.83409
#05 multilearn T	0.38258	0.70709	0.80200	0.37579	0.67780	0.83182
#06 multilearn+forgets T	0.37654	0.71762	0.80200	0.35304	0.74304	0.83182
#07 multigs+multilearn T	0.38117	0.72406	0.80900	0.37772	0.68400	0.83409
#08 multigs+multilearn+forgets T	0.37325	0.73186	0.80900	0.35239	0.74876	0.83409
#09 multiprior T	0.38276	0.66608	0.80316	0.37208	0.67767	0.83732
#10 multiprior+forgets T	0.38141	0.67476	0.80105	0.35783	0.70889	0.83732
#11 multigs TE	0.37838	0.73265	0.82200	0.36882	0.71893	0.84545
#12 multigs+forgets TE	0.36964	0.74111	0.82200	0.34577	0.77066	0.84545
#13 multilearn TE	0.38256	0.70756	0.80200	0.37578	0.67830	0.83182
#14 multilearn+forgets TE	0.37643	0.71853	0.80200	0.35304	0.74214	0.83182
#15 multigs+multilearn TE	0.37804	0.73338	0.82200	0.36816	0.72103	0.84545
#16 multigs+multilearn+forgets TE	0.36859	0.74356	0.82300	0.34472	0.77073	0.84318
#17 multiprior TE	0.38276	0.66608	0.80316	0.37208	0.67767	0.83732
#18 multiprior+forgets TE	0.38141	0.67476	0.80105	0.35783	0.70889	0.83732

Table B5. Student performance prediction – DKC 05.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.36151	0.86778	0.79600	0.37534	0.86704	0.78478
#02 vanilla+forgets	0.35752	0.87665	0.80000	0.37706	0.87666	0.77391
#03 multigs T	0.35240	0.87858	0.79500	0.36783	0.87670	0.80870
#04 multigs+forgets T	0.34945	0.88740	0.79800	0.36837	0.87979	0.78043
#05 multilearn T	0.36176	0.86813	0.79600	0.37405	0.86858	0.78696
#06 multilearn+forgets T	0.35816	0.87294	0.80300	0.37435	0.87140	0.78261
#07 multigs+multilearn T	0.35626	0.87993	0.78500	0.36697	0.87520	0.78913
#08 multigs+multilearn+forgets T	0.34756	0.88779	0.80100	0.37106	0.88113	0.80217
#09 multiprior T	0.36678	0.83801	0.79789	0.37743	0.85352	0.78261
#10 multiprior+forgets T	0.35803	0.84890	0.80316	0.37259	0.86268	0.78719
#11 multigs TE	0.35529	0.87845	0.77800	0.35356	0.88240	0.79348
#12 multigs+forgets TE	0.35025	0.88826	0.80200	0.35846	0.88338	0.78261
#13 multilearn TE	0.36164	0.86777	0.79600	0.37533	0.86712	0.78478
#14 multilearn+forgets TE	0.35767	0.86931	0.80100	0.37666	0.87682	0.77391
#15 multigs+multilearn TE	0.35507	0.87958	0.77500	0.35428	0.88190	0.79130
#16 multigs+multilearn+forgets TE	0.35039	0.88381	0.79800	0.35874	0.88178	0.78261
#17 multiprior TE	0.36639	0.83802	0.79789	0.37743	0.85352	0.78261
#18 multiprior+forgets TE	0.35720	0.84813	0.80316	0.37259	0.86268	0.78719

Table B6. Student performance prediction – DKC 06.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.21166	0.83492	0.91333	0.28182	0.87037	0.85000
#02 vanilla+forgets	0.21193	0.82837	0.91222	0.27597	0.86294	0.86818
#03 multigs T	0.20052	0.84718	0.94333	0.28316	0.87056	0.85227
#04 multigs+forgets T	0.20109	0.84851	0.94556	0.27680	0.86644	0.86591
#05 multilearn T	0.21661	0.83350	0.91000	0.28111	0.87222	0.85000
#06 multilearn+forgets T	0.21301	0.82470	0.91333	0.27790	0.86481	0.86364
#07 multigs+multilearn T	0.20055	0.84702	0.94333	0.28307	0.87056	0.85227
#08 multigs+multilearn+forgets T	0.20107	0.84797	0.94333	0.27710	0.86681	0.86136
#09 multiprior T	0.21113	0.79187	0.91345	0.27826	0.82358	0.86364
#10 multiprior+forgets T	0.20949	0.79272	0.91579	0.27408	0.82379	0.88517
#11 multigs TE	0.21254	0.83469	0.91333	0.28019	0.87056	0.86364
#12 multigs+forgets TE	0.21352	0.82813	0.91222	0.27315	0.86656	0.88864
#13 multilearn TE	0.21661	0.83382	0.91000	0.28111	0.87219	0.85000
#14 multilearn+forgets TE	0.21267	0.82934	0.91556	0.27788	0.86312	0.86364
#15 multigs+multilearn TE	0.21244	0.83437	0.91222	0.27977	0.87075	0.87500
#16 multigs+multilearn+forgets TE	0.21288	0.82876	0.91667	0.27367	0.86669	0.88636
#17 multiprior TE	0.21113	0.79187	0.91345	0.27826	0.82358	0.86364
#18 multiprior+forgets TE	0.20949	0.79272	0.91579	0.27408	0.82379	0.88517

Table B7. Student performance prediction – DKC 07.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.41772	0.80673	0.75105	0.43582	0.78758	0.72738
#02 vanilla+forgets	0.41800	0.80718	0.75158	0.43636	0.78253	0.72262
#03 multigs T	0.42291	0.80781	0.73211	0.40661	0.83528	0.76786
#04 multigs+forgets T	0.42386	0.82564	0.73105	0.41473	0.83168	0.75357
#05 multilearn T	0.41757	0.80407	0.75316	0.43712	0.78651	0.72976
#06 multilearn+forgets T	0.42071	0.80131	0.74579	0.43663	0.78260	0.71548
#07 multigs+multilearn T	0.42259	0.80805	0.73263	0.40631	0.83514	0.76786
#08 multigs+multilearn+forgets T	0.41849	0.82150	0.73474	0.40938	0.83319	0.76190
#09 multiprior T	0.42624	0.79207	0.72521	0.44690	0.77425	0.70050
#10 multiprior+forgets T	0.42532	0.79296	0.71967	0.44563	0.77157	0.69048
#11 multigs TE	0.41813	0.81479	0.74632	0.40773	0.83536	0.76905
#12 multigs+forgets TE	0.42137	0.83271	0.74158	0.41439	0.83421	0.74286
#13 multilearn TE	0.41767	0.80697	0.75053	0.43703	0.78809	0.72976
#14 multilearn+forgets TE	0.41946	0.80457	0.74474	0.43637	0.78396	0.71786
#15 multigs+multilearn TE	0.41789	0.81544	0.74789	0.40745	0.83481	0.76786
#16 multigs+multilearn+forgets TE	0.41116	0.83305	0.75053	0.40791	0.83326	0.76190
#17 multiprior TE	0.42571	0.79339	0.72742	0.44705	0.77367	0.70050
#18 multiprior+forgets TE	0.42464	0.79438	0.72022	0.44577	0.77122	0.69048

Table B8. Student performance prediction – DKC 08.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.37660	0.74622	0.78556	0.40262	0.75545	0.74872
#02 vanilla+forgets	0.36063	0.77813	0.81056	0.36754	0.83745	0.80769
#03 multigs T	0.37166	0.78234	0.82944	0.39333	0.80042	0.82821
#04 multigs+forgets T	0.36469	0.78503	0.82000	0.36911	0.85191	0.80769
#05 multilearn T	0.37654	0.74627	0.78556	0.40257	0.75530	0.74872
#06 multilearn+forgets T	0.36036	0.77627	0.81222	0.36764	0.83646	0.80769
#07 multigs+multilearn T	0.37167	0.78222	0.82944	0.39333	0.80045	0.82821
#08 multigs+multilearn+forgets T	0.35327	0.80595	0.82167	0.35450	0.86381	0.81795
#09 multiprior T	0.37943	0.69329	0.79942	0.41524	0.74679	0.73819
#10 multiprior+forgets T	0.36490	0.71566	0.81813	0.37392	0.80815	0.80027
#11 multigs TE	0.36193	0.79223	0.84833	0.38083	0.83183	0.84615
#12 multigs+forgets TE	0.33872	0.81857	0.85611	0.33832	0.88142	0.85769
#13 multilearn TE	0.37654	0.74659	0.78556	0.40257	0.75549	0.74872
#14 multilearn+forgets TE	0.36012	0.77884	0.81444	0.36710	0.83598	0.80769
#15 multigs+multilearn TE	0.36190	0.79233	0.84833	0.38080	0.83186	0.84615
#16 multigs+multilearn+forgets TE	0.34582	0.80610	0.84667	0.34368	0.87771	0.84103
#17 multiprior TE	0.37807	0.69719	0.79942	0.41403	0.75066	0.73819
#18 multiprior+forgets TE	0.36256	0.71930	0.81813	0.37075	0.81169	0.80027

Table B9. Student performance prediction – DKC 09.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.44427	0.68172	0.66842	0.43204	0.74487	0.65000
#02 vanilla+forgets	0.42325	0.76789	0.69789	0.42530	0.77407	0.68690
#03 multigs T	0.44160	0.70276	0.71316	0.42239	0.77525	0.73452
#04 multigs+forgets T	0.41858	0.77961	0.70105	0.41456	0.80290	0.71310
#05 multilearn T	0.44422	0.68171	0.66842	0.43203	0.74481	0.65000
#06 multilearn+forgets T	0.42396	0.76389	0.69000	0.42413	0.77355	0.68810
#07 multigs+multilearn T	0.44162	0.70233	0.71263	0.42293	0.77523	0.73452
#08 multigs+multilearn+forgets T	0.41944	0.78365	0.70842	0.41396	0.80812	0.71905
#09 multiprior T	0.45421	0.68596	0.68144	0.44802	0.74228	0.68922
#10 multiprior+forgets T	0.43325	0.73382	0.70637	0.43464	0.75248	0.69424
#11 multigs TE	0.39993	0.82721	0.84158	0.38519	0.84825	0.81310
#12 multigs+forgets TE	0.37322	0.85723	0.83789	0.37557	0.85991	0.80714
#13 multilearn TE	0.44421	0.68188	0.66842	0.43202	0.74504	0.65000
#14 multilearn+forgets TE	0.42308	0.76972	0.69789	0.42594	0.77533	0.68690
#15 multigs+multilearn TE	0.39993	0.82715	0.84158	0.38520	0.84827	0.81310
#16 multigs+multilearn+forgets TE	0.37498	0.85015	0.81684	0.38308	0.85208	0.80476
#17 multiprior TE	0.45299	0.69332	0.68144	0.44816	0.74493	0.68797
#18 multiprior+forgets TE	0.42999	0.74197	0.71080	0.43411	0.75479	0.69298

Table B10. Student performance prediction – DKC 10.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.40986	0.81448	0.76824	0.39840	0.83525	0.77821
#02 vanilla+forgets	0.40693	0.82844	0.76647	0.39893	0.84100	0.77436
#03 multigs T	0.40370	0.82334	0.76176	0.40350	0.83160	0.77051
#04 multigs+forgets T	0.39577	0.85339	0.78941	0.39197	0.86059	0.79615
#05 multilearn T	0.40939	0.81463	0.76941	0.39770	0.83465	0.77821
#06 multilearn+forgets T	0.40785	0.82897	0.76176	0.39926	0.84045	0.76795
#07 multigs+multilearn T	0.39773	0.84297	0.78765	0.39545	0.85312	0.78718
#08 multigs+multilearn+forgets T	0.39123	0.85163	0.78941	0.38723	0.85305	0.79615
#09 multiprior T	0.42562	0.80355	0.74118	0.41231	0.82222	0.74224
#10 multiprior+forgets T	0.41721	0.81478	0.73808	0.41171	0.82580	0.75169
#11 multigs TE	0.38159	0.86719	0.80941	0.37560	0.87037	0.81538
#12 multigs+forgets TE	0.38119	0.87503	0.81412	0.37498	0.87873	0.81667
#13 multilearn TE	0.40936	0.81358	0.76824	0.39754	0.83549	0.77821
#14 multilearn+forgets TE	0.40721	0.82427	0.76412	0.39820	0.84354	0.77436
#15 multigs+multilearn TE	0.38152	0.86710	0.80941	0.37551	0.87045	0.81538
#16 multigs+multilearn+forgets TE	0.37878	0.86871	0.81588	0.37188	0.87779	0.81795
#17 multiprior TE	0.42362	0.80662	0.75666	0.40952	0.82449	0.76248
#18 multiprior+forgets TE	0.41468	0.81910	0.75604	0.40885	0.82855	0.75574

Table B11. Student performance prediction – DKC 11.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.36904	0.85398	0.78647	0.36968	0.87817	0.80897
#02 vanilla+forgets	0.34957	0.87895	0.81471	0.35443	0.90427	0.82821
#03 multigs T	0.34898	0.88184	0.85706	0.36818	0.88301	0.82564
#04 multigs+forgets T	0.32732	0.90555	0.85412	0.34689	0.91067	0.81795
#05 multilearn T	0.36859	0.85416	0.78765	0.36936	0.87831	0.80897
#06 multilearn+forgets T	0.35016	0.87422	0.81529	0.35503	0.89795	0.82821
#07 multigs+multilearn T	0.34962	0.88143	0.85765	0.36900	0.88349	0.82821
#08 multigs+multilearn+forgets T	0.32681	0.90573	0.85412	0.34621	0.90991	0.81795
#09 multiprior T	0.37885	0.81773	0.78700	0.37745	0.84998	0.81107
#10 multiprior+forgets T	0.35475	0.84293	0.81424	0.35833	0.87194	0.82726
#11 multigs TE	0.34548	0.88807	0.86000	0.33919	0.91722	0.88974
#12 multigs+forgets TE	0.32198	0.90882	0.86235	0.31583	0.93466	0.89487
#13 multilearn TE	0.36864	0.85401	0.78765	0.36938	0.87816	0.80897
#14 multilearn+forgets TE	0.34975	0.87605	0.81353	0.35464	0.90307	0.82821
#15 multigs+multilearn TE	0.34566	0.88803	0.86059	0.33933	0.91719	0.88846
#16 multigs+multilearn+forgets TE	0.32167	0.90730	0.86176	0.31491	0.93389	0.89231
#17 multiprior TE	0.37814	0.81819	0.78700	0.37489	0.85102	0.81107
#18 multiprior+forgets TE	0.35392	0.84339	0.81610	0.35485	0.87416	0.83401

Table B12. Student performance prediction – DKC 12.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.42138	0.80291	0.73312	0.39488	0.85265	0.78611
#02 vanilla+forgets	0.40478	0.82717	0.76438	0.38355	0.86394	0.80694
#03 multigs T	0.41434	0.80974	0.77375	0.40120	0.85872	0.79722
#04 multigs+forgets T	0.39595	0.84989	0.76688	0.38456	0.88440	0.78056
#05 multilearn T	0.42194	0.80221	0.73562	0.39461	0.85397	0.78472
#06 multilearn+forgets T	0.40516	0.82604	0.76312	0.38260	0.86175	0.80694
#07 multigs+multilearn T	0.41424	0.80961	0.77500	0.40096	0.85907	0.79861
#08 multigs+multilearn+forgets T	0.39469	0.84560	0.76688	0.38360	0.88114	0.77778
#09 multiprior T	0.42449	0.79579	0.74605	0.40704	0.83041	0.76754
#10 multiprior+forgets T	0.40595	0.81008	0.76447	0.39413	0.83665	0.78216
#11 multigs TE	0.39090	0.86140	0.83562	0.38188	0.90391	0.84028
#12 multigs+forgets TE	0.37077	0.88191	0.83438	0.36038	0.90456	0.84583
#13 multilearn TE	0.42147	0.80462	0.73562	0.39466	0.85353	0.78750
#14 multilearn+forgets TE	0.40464	0.82670	0.76688	0.38403	0.86405	0.80417
#15 multigs+multilearn TE	0.38961	0.86136	0.83438	0.37962	0.90518	0.84306
#16 multigs+multilearn+forgets TE	0.36421	0.87915	0.83375	0.35370	0.90554	0.84444
#17 multiprior TE	0.42274	0.80223	0.74605	0.40473	0.83297	0.76754
#18 multiprior+forgets TE	0.40342	0.81368	0.76447	0.39033	0.84032	0.79094

Table B13. Student performance prediction – DKC 13.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.39137	0.81420	0.77000	0.41582	0.80775	0.74459
#02 vanilla+forgets	0.37109	0.84601	0.80471	0.40114	0.83275	0.76757
#03 multigs T	0.38851	0.81325	0.77235	0.41927	0.80419	0.74054
#04 multigs+forgets T	0.36936	0.84764	0.80529	0.40237	0.83117	0.76757
#05 multilearn T	0.39084	0.81464	0.77118	0.41581	0.80608	0.74459
#06 multilearn+forgets T	0.37141	0.84579	0.80353	0.40103	0.83330	0.76757
#07 multigs+multilearn T	0.38765	0.81367	0.77529	0.41718	0.80642	0.74459
#08 multigs+multilearn+forgets T	0.36929	0.84742	0.80471	0.40227	0.83176	0.76757
#09 multiprior T	0.40237	0.79318	0.76842	0.42899	0.77602	0.72262
#10 multiprior+forgets T	0.37683	0.82062	0.80433	0.40650	0.80603	0.76102
#11 multigs TE	0.36342	0.87240	0.85529	0.38920	0.85660	0.83378
#12 multigs+forgets TE	0.35057	0.89295	0.85059	0.37920	0.87700	0.82432
#13 multilearn TE	0.39086	0.81431	0.77118	0.41592	0.80639	0.74459
#14 multilearn+forgets TE	0.37077	0.84707	0.80647	0.40033	0.83536	0.77297
#15 multigs+multilearn TE	0.36331	0.87216	0.85471	0.38906	0.85671	0.83378
#16 multigs+multilearn+forgets TE	0.34303	0.89466	0.86294	0.37129	0.87744	0.83784
#17 multiprior TE	0.39925	0.79804	0.76842	0.42703	0.78088	0.72262
#18 multiprior+forgets TE	0.37224	0.82632	0.80743	0.40401	0.80962	0.76102

Table B14. Student performance prediction – DKC 14.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.42461	0.78888	0.72250	0.41951	0.79930	0.74143
#02 vanilla+forgets	0.40030	0.83277	0.76938	0.39353	0.84879	0.78571
#03 multigs T	0.40653	0.84067	0.78188	0.40132	0.85890	0.78571
#04 multigs+forgets T	0.39173	0.86804	0.79875	0.38564	0.88788	0.81143
#05 multilearn T	0.42354	0.78985	0.72562	0.41858	0.80064	0.75000
#06 multilearn+forgets T	0.40029	0.83199	0.77188	0.39389	0.84615	0.78571
#07 multigs+multilearn T	0.40672	0.84718	0.78812	0.40152	0.86413	0.79000
#08 multigs+multilearn+forgets T	0.38936	0.86375	0.79812	0.38240	0.88775	0.81429
#09 multiprior T	0.43288	0.77459	0.72105	0.43647	0.78876	0.71579
#10 multiprior+forgets T	0.40761	0.81600	0.75000	0.40433	0.83935	0.76391
#11 multigs TE	0.38062	0.86947	0.80188	0.36903	0.88807	0.83571
#12 multigs+forgets TE	0.36838	0.88880	0.81500	0.35093	0.91383	0.85000
#13 multilearn TE	0.42416	0.78942	0.72438	0.41857	0.80147	0.75000
#14 multilearn+forgets TE	0.40101	0.83545	0.76000	0.39463	0.84736	0.78571
#15 multigs+multilearn TE	0.38086	0.86889	0.80125	0.36922	0.88768	0.83286
#16 multigs+multilearn+forgets TE	0.36570	0.88354	0.81438	0.34617	0.90999	0.85000
#17 multiprior TE	0.43055	0.77693	0.73026	0.43435	0.79027	0.71579
#18 multiprior+forgets TE	0.40441	0.81883	0.76382	0.40151	0.84061	0.78045

Table B15. Student performance prediction – DKC 15.

BKT model	Cross-validation			Evaluation		
	RMSE	**AUC**	**Accuracy**	**RMSE**	**AUC**	**Accuracy**
#01 vanilla	0.36693	0.88152	0.81750	0.39103	0.86347	0.78243
#02 vanilla+forgets	0.36369	0.89147	0.81500	0.37707	0.87951	0.79459
#03 multigs T	0.35771	0.89746	0.83375	0.36828	0.89017	0.84189
#04 multigs+forgets T	0.35466	0.91109	0.82750	0.36258	0.90702	0.82162
#05 multilearn T	0.36640	0.88032	0.81875	0.38677	0.86886	0.78784
#06 multilearn+forgets T	0.36315	0.88882	0.81750	0.37841	0.87710	0.79189
#07 multigs+multilearn T	0.35782	0.89733	0.83375	0.36863	0.89027	0.84189
#08 multigs+multilearn+forgets T	0.35026	0.90734	0.82687	0.35777	0.90318	0.82297
#09 multiprior T	0.38011	0.87547	0.79868	0.40001	0.85421	0.76956
#10 multiprior+forgets T	0.36853	0.88699	0.81842	0.38214	0.86554	0.80228
#11 multigs TE	0.31436	0.93542	0.88250	0.32648	0.93103	0.88649
#12 multigs+forgets TE	0.31233	0.94020	0.88188	0.32170	0.93661	0.87162
#13 multilearn TE	0.36627	0.88135	0.81688	0.38685	0.86892	0.78649
#14 multilearn+forgets TE	0.36299	0.89115	0.81687	0.37670	0.87692	0.79595
#15 multigs+multilearn TE	0.31522	0.93386	0.87875	0.32754	0.92982	0.88649
#16 multigs+multilearn+forgets TE	0.30962	0.93702	0.87875	0.31721	0.92936	0.87297
#17 multiprior TE	0.37543	0.87762	0.81776	0.39430	0.85881	0.78947
#18 multiprior+forgets TE	0.36359	0.89060	0.82237	0.37551	0.87006	0.82219

Table B16. Student performance prediction – DKC 16.

BKT model	Cross-validation			Evaluation		
	RMSE	**AUC**	**Accuracy**	**RMSE**	**AUC**	**Accuracy**
#01 vanilla	0.33157	0.88507	0.85375	0.35756	0.86421	0.82973
#02 vanilla+forgets	0.30980	0.89896	0.87188	0.32813	0.88650	0.84459
#03 multigs T	0.32836	0.87715	0.87500	0.35259	0.85339	0.84595
#04 multigs+forgets T	0.30442	0.90925	0.87063	0.31742	0.89977	0.84865
#05 multilearn T	0.32889	0.88756	0.86000	0.35728	0.86291	0.83243
#06 multilearn+forgets T	0.31074	0.89869	0.87188	0.32583	0.88503	0.85541
#07 multigs+multilearn T	0.32791	0.87731	0.87500	0.35229	0.85374	0.84459
#08 multigs+multilearn+forgets T	0.30109	0.91039	0.87188	0.32165	0.89534	0.85135
#09 multiprior T	0.32043	0.85130	0.87500	0.35276	0.84789	0.84068
#10 multiprior+forgets T	0.30514	0.86181	0.87566	0.33013	0.86059	0.85064
#11 multigs TE	0.32144	0.90187	0.90500	0.33870	0.89167	0.87973
#12 multigs+forgets TE	0.29648	0.91748	0.90250	0.31432	0.91175	0.88378
#13 multilearn TE	0.32890	0.88770	0.86000	0.35732	0.85980	0.83243
#14 multilearn+forgets TE	0.31165	0.89893	0.87250	0.32689	0.88817	0.85405
#15 multigs+multilearn TE	0.32138	0.90177	0.90500	0.33865	0.89160	0.87973
#16 multigs+multilearn+forgets TE	0.29133	0.92126	0.90250	0.30963	0.91690	0.88243
#17 multiprior TE	0.31838	0.85640	0.87500	0.34832	0.85343	0.84068
#18 multiprior+forgets TE	0.30102	0.86560	0.87632	0.32553	0.86349	0.85064

Table B17. Student performance prediction – DKC 17.

BKT model	Cross-validation			Evaluation		
	RMSE	**AUC**	**Accuracy**	**RMSE**	**AUC**	**Accuracy**
#01 vanilla	0.38899	0.82836	0.80786	0.37133	0.85356	0.81818
#02 vanilla+forgets	0.37917	0.84892	0.81571	0.35537	0.87700	0.82879
#03 multigs T	0.39306	0.86245	0.82071	0.37285	0.86681	0.84697
#04 multigs+forgets T	0.38155	0.87815	0.83286	0.36258	0.89508	0.85000
#05 multilearn T	0.38904	0.82846	0.80786	0.37178	0.85377	0.81667
#06 multilearn+forgets T	0.38283	0.85158	0.80143	0.35731	0.87874	0.83030
#07 multigs+multilearn T	0.39281	0.86200	0.81857	0.37274	0.86673	0.84545
#08 multigs+multilearn+forgets T	0.37775	0.86003	0.83429	0.35598	0.87118	0.85152
#09 multiprior T	0.41394	0.81471	0.75489	0.38720	0.85320	0.77831
#10 multiprior+forgets T	0.40523	0.82386	0.76391	0.37486	0.86352	0.80702
#11 multigs TE	0.34926	0.89690	0.86714	0.33672	0.90751	0.87576
#12 multigs+forgets TE	0.34451	0.90930	0.86786	0.32701	0.93133	0.87576
#13 multilearn TE	0.38898	0.82858	0.80929	0.37154	0.85373	0.81667
#14 multilearn+forgets TE	0.38820	0.85003	0.79357	0.35537	0.87748	0.83485
#15 multigs+multilearn TE	0.34917	0.89693	0.86714	0.33665	0.90747	0.87576
#16 multigs+multilearn+forgets TE	0.33826	0.91246	0.86643	0.31762	0.93412	0.87727
#17 multiprior TE	0.41136	0.81883	0.76617	0.38446	0.85723	0.78309
#18 multiprior+forgets TE	0.40124	0.82640	0.76992	0.37112	0.86871	0.81180

Table B18. Student performance prediction – DKC 19.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.36717	0.84659	0.80200	0.35975	0.87743	0.80294
#02 vanilla+forgets	0.32135	0.89613	0.85667	0.31127	0.92909	0.87059
#03 multigs T	0.35391	0.85684	0.83133	0.35618	0.87234	0.80294
#04 multigs+forgets T	0.31893	0.91181	0.83200	0.30550	0.93578	0.88382
#05 multilearn T	0.36717	0.84846	0.80000	0.35954	0.87746	0.80294
#06 multilearn+forgets T	0.32168	0.89800	0.86200	0.31167	0.92895	0.86912
#07 multigs+multilearn T	0.35402	0.85691	0.83133	0.35625	0.87235	0.80294
#08 multigs+multilearn+forgets T	0.31842	0.91111	0.84533	0.31433	0.93483	0.88088
#09 multiprior T	0.36901	0.84043	0.80842	0.36887	0.86914	0.81424
#10 multiprior+forgets T	0.32999	0.87283	0.84070	0.31566	0.91206	0.84830
#11 multigs TE	0.32525	0.89113	0.88600	0.31827	0.91896	0.87647
#12 multigs+forgets TE	0.28564	0.93336	0.90533	0.27292	0.96009	0.90294
#13 multilearn TE	0.36718	0.84681	0.80000	0.35955	0.87748	0.80294
#14 multilearn+forgets TE	0.32140	0.89644	0.85133	0.31042	0.93212	0.86618
#15 multigs+multilearn TE	0.32538	0.89104	0.88400	0.31842	0.91880	0.87059
#16 multigs+multilearn+forgets TE	0.28260	0.93170	0.90400	0.26901	0.96121	0.90441
#17 multiprior TE	0.36655	0.84274	0.80842	0.36494	0.87528	0.81424
#18 multiprior+forgets TE	0.32698	0.87651	0.84491	0.30855	0.91422	0.85913

Table B19. Student performance prediction – DKC 21.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.34348	0.88215	0.85200	0.37297	0.89395	0.81471
#02 vanilla+forgets	0.32469	0.92676	0.86200	0.36013	0.91169	0.82500
#03 multigs T	0.30720	0.93876	0.87867	0.35304	0.90114	0.85588
#04 multigs+forgets T	0.30375	0.94991	0.88600	0.34327	0.91481	0.84265
#05 multilearn T	0.34363	0.88230	0.85133	0.37451	0.89462	0.81029
#06 multilearn+forgets T	0.32652	0.92683	0.85467	0.36207	0.90821	0.81912
#07 multigs+multilearn T	0.30720	0.93874	0.87867	0.35304	0.90106	0.85588
#08 multigs+multilearn+forgets T	0.29685	0.94578	0.88533	0.33301	0.90875	0.84706
#09 multiprior T	0.36498	0.87654	0.83789	0.38497	0.87724	0.80495
#10 multiprior+forgets T	0.34642	0.91312	0.83649	0.37303	0.89178	0.80960
#11 multigs TE	0.29898	0.93863	0.88133	0.33590	0.91739	0.86618
#12 multigs+forgets TE	0.29597	0.94908	0.89467	0.32716	0.93409	0.85882
#13 multilearn TE	0.34351	0.88196	0.85267	0.37471	0.89199	0.81176
#14 multilearn+forgets TE	0.32617	0.92677	0.85733	0.36238	0.91210	0.82353
#15 multigs+multilearn TE	0.29898	0.93861	0.88133	0.33590	0.91733	0.86618
#16 multigs+multilearn+forgets TE	0.29088	0.94639	0.89133	0.32242	0.93208	0.85735
#17 multiprior TE	0.36040	0.87835	0.84211	0.38348	0.88081	0.80650
#18 multiprior+forgets TE	0.34177	0.91003	0.84070	0.37143	0.89596	0.81115

Table B20. Student performance prediction – DKC 23.

BKT model	Cross-validation			Evaluation		
	RMSE	AUC	Accuracy	RMSE	AUC	Accuracy
#01 vanilla	0.39917	0.82490	0.78154	0.40703	0.81877	0.76552
#02 vanilla+forgets	0.39504	0.84393	0.78308	0.38342	0.85563	0.79483
#03 multigs T	0.36304	0.88809	0.83923	0.39186	0.85649	0.78793
#04 multigs+forgets T	0.35805	0.89983	0.84385	0.38081	0.87341	0.79310
#05 multilearn T	0.39904	0.82501	0.78231	0.40698	0.81989	0.76724
#06 multilearn+forgets T	0.39468	0.84363	0.77923	0.38311	0.85270	0.79483
#07 multigs+multilearn T	0.36285	0.88832	0.84000	0.39174	0.85663	0.78793
#08 multigs+multilearn+forgets T	0.35570	0.90214	0.84692	0.37525	0.87434	0.80000
#09 multiprior T	0.41038	0.80857	0.76113	0.41507	0.80425	0.75681
#10 multiprior+forgets T	0.40390	0.82671	0.77166	0.39440	0.83605	0.78766
#11 multigs TE	0.34890	0.89283	0.85769	0.36413	0.88406	0.82414
#12 multigs+forgets TE	0.34433	0.90393	0.86308	0.35555	0.89501	0.83621
#13 multilearn TE	0.39920	0.82490	0.78231	0.40705	0.81715	0.76552
#14 multilearn+forgets TE	0.39504	0.84130	0.78385	0.38269	0.85902	0.79483
#15 multigs+multilearn TE	0.34892	0.89281	0.85769	0.36420	0.88432	0.82414
#16 multigs+multilearn+forgets TE	0.34098	0.90526	0.86462	0.35113	0.89811	0.83793
#17 multiprior TE	0.40825	0.80979	0.77004	0.41364	0.80624	0.76407
#18 multiprior+forgets TE	0.40156	0.82812	0.77976	0.39278	0.83729	0.79492

Table B21. Student performance prediction – DKC 25.

APPENDIX C

The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 01-DKC 25

The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 01-DKC 25

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.84683	0.00000	0.42945	0.00149	0.27976	0.08040	0.85714	0.68852	0.63158
#02 vanilla+forgets	0.83760	0.00000	0.44832	0.00086	0.26620	0.09684	0.87273	0.73333	0.60714
#03 multigs T	0.72752	0.00000	0.40484	0.00291	0.21684	0.17896	0.69565	0.70270	0.62857
#04 multigs+forgets T	0.77271	0.00000	0.45910	0.00062	0.29212	0.06739	0.00000	0.00000	0.00000
#05 multilearn T	0.84407	0.00000	0.42640	0.00162	0.29195	0.06756	0.85714	0.68852	0.63158
#06 multilearn+forgets T	0.84727	0.00000	0.45537	0.00069	0.27638	0.08429	0.92308	0.73684	0.60377
#07 multigs+multilearn T	0.72909	0.00000	0.40348	0.00302	0.21842	0.17573	0.69565	0.70270	0.62857
#08 multigs+multilearn+forgets T	0.77017	0.00000	0.46526	0.00051	0.28193	0.07798	0.00000	0.00000	0.00000
#09 multiprior T	0.84312	0.00000	0.44581	0.00093	0.26354	0.10036	0.88889	0.74576	0.61818
#10 multiprior+forgets T	0.73700	0.00000	0.45182	0.00077	0.32277	0.04222	0.00000	0.00000	0.00000
#11 multigs TE	0.72654	0.00000	0.40519	0.00288	0.21586	0.18098	0.69565	0.70270	0.62857
#12 multigs+forgets TE	0.72931	0.00000	0.40954	0.00257	0.21562	0.18146	0.69565	0.70270	0.62857
#13 multilearn TE	0.84358	0.00000	0.42637	0.00162	0.29144	0.06806	0.85714	0.68852	0.63158
#14 multilearn+forgets TE	0.84592	0.00000	0.45519	0.00070	0.27542	0.08542	0.92308	0.73684	0.60377
#15 multigs+multilearn TE	0.72938	0.00000	0.40380	0.00299	0.21861	0.17534	0.69565	0.70270	0.62857
#16 multigs+multilearn+forgets TE	0.75350	0.00000	0.46473	0.00052	0.27709	0.08345	0.00000	0.00000	0.00000
#17 multiprior TE	0.84312	0.00000	0.44581	0.00093	0.26354	0.10036	0.88889	0.74576	0.61818
#18 multiprior+forgets TE	0.73700	0.00000	0.45182	0.00077	0.32277	0.04222	0.00000	0.00000	0.00000

Table C1. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 01.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.86570	0.00000	0.27551	0.03805	0.39495	0.00725	0.96296	0.69333	0.70423
#02 vanilla+forgets	0.85405	0.00000	0.19369	0.14884	0.31595	0.03449	0.98734	0.65753	0.66667
#03 multigs T	0.55802	0.00000	0.18453	0.16940	0.09065	0.55372	0.76471	0.62500	0.56522
#04 multigs+forgets T	0.44608	0.00009	0.08440	0.53250	0.00987	0.94870	0.73585	0.62000	0.56250
#05 multilearn T	0.87649	0.00000	0.24757	0.06335	0.33495	0.02451	0.96104	0.67606	0.68657
#06 multilearn+forgets T	0.87842	0.00000	0.26800	0.04385	0.33776	0.02326	0.93333	0.66667	0.70769
#07 multigs+multilearn T	0.55875	0.00000	0.18457	0.16930	0.09131	0.55080	0.76471	0.62500	0.56522
#08 multigs+multilearn+forgets T	0.50549	0.00001	0.13786	0.30649	-0.00255	0.98674	0.75000	0.61224	0.55319
#09 multiprior T	0.86690	0.00000	0.21625	0.10617	0.32021	0.03200	0.97436	0.66667	0.67647
#10 multiprior+forgets T	0.76652	0.00000	0.24053	0.07150	0.36575	0.01348	0.00000	0.00000	0.00000
#11 multigs TE	0.54984	0.00000	0.18651	0.16478	0.09326	0.54229	0.76471	0.62500	0.56522
#12 multigs+forgets TE	0.44402	0.00009	0.08388	0.53501	0.00872	0.95465	0.73585	0.62000	0.56250
#13 multilearn TE	0.87124	0.00000	0.23924	0.07308	0.35016	0.01837	0.96203	0.68493	0.69565
#14 multilearn+forgets TE	0.87154	0.00000	0.22016	0.09984	0.32764	0.02801	0.97436	0.66667	0.67647
#15 multigs+multilearn TE	0.55566	0.00000	0.18328	0.17236	0.08817	0.56465	0.76471	0.62500	0.56522
#16 multigs+multilearn+forgets TE	0.51550	0.00000	0.14573	0.27942	0.00388	0.97983	0.76471	0.62500	0.56522
#17 multiprior TE	0.86690	0.00000	0.21625	0.10617	0.32021	0.03200	0.97436	0.66667	0.67647
#18 multiprior+forgets TE	0.76652	0.00000	0.24053	0.07150	0.36575	0.01348	0.00000	0.00000	0.00000

Table C2. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 02.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.81217	0.00000	0.57394	0.00002	0.40079	0.00853	0.58182	0.83333	0.76471
#02 vanilla+forgets	0.71139	0.00000	0.35348	0.01272	0.32313	0.03686	0.00000	0.00000	0.00000
#03 multigs T	0.68882	0.00000	0.44202	0.00147	0.43368	0.00412	0.51613	0.78481	0.74667
#04 multigs+forgets T	0.62772	0.00000	0.29633	0.03869	0.36948	0.01603	0.00000	0.00000	0.00000
#05 multilearn T	0.81967	0.00000	0.60234	0.00000	0.41561	0.00620	0.59259	0.84507	0.77612
#06 multilearn+forgets T	0.71738	0.00000	0.35742	0.01169	0.31484	0.04228	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.68885	0.00000	0.44143	0.00150	0.43369	0.00411	0.51613	0.78481	0.74667
#08 multigs+multilearn+forgets T	0.63151	0.00000	0.30488	0.03317	0.36293	0.01816	0.00000	0.00000	0.00000
#09 multiprior T	0.86383	0.00000	0.57222	0.00002	0.40919	0.00713	0.81081	0.66667	0.64000
#10 multiprior+forgets T	0.62643	0.00000	0.29914	0.03680	0.27346	0.07972	0.00000	0.00000	0.00000
#11 multigs TE	0.69667	0.00000	0.45118	0.00114	0.42109	0.00549	0.52459	0.79487	0.72973
#12 multigs+forgets TE	0.69093	0.00000	0.34537	0.01508	0.35428	0.02134	0.00000	0.00000	0.00000
#13 multilearn TE	0.81225	0.00000	0.57474	0.00002	0.40026	0.00863	0.58182	0.83333	0.76471
#14 multilearn+forgets TE	0.70435	0.00000	0.34972	0.01377	0.32362	0.03656	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.69789	0.00000	0.45435	0.00104	0.42500	0.00502	0.52459	0.79487	0.72973
#16 multigs+multilearn+forgets TE	0.64712	0.00000	0.31644	0.02675	0.35333	0.02171	0.00000	0.00000	0.00000
#17 multiprior TE	0.86383	0.00000	0.57222	0.00002	0.40919	0.00713	0.81081	0.66667	0.64000
#18 multiprior+forgets TE	0.62643	0.00000	0.29914	0.03680	0.27346	0.07972	0.00000	0.00000	0.00000

Table C3. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 03.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.84782	0.00000	0.55356	0.00014	0.40084	0.02544	0.87273	0.73077	0.69231
#02 vanilla+forgets	0.84379	0.00000	0.54668	0.00018	0.38648	0.03174	0.87273	0.73077	0.69231
#03 multigs T	0.75685	0.00000	0.42422	0.00511	0.29678	0.10496	0.80000	0.70175	0.66667
#04 multigs+forgets T	0.79509	0.00000	0.38788	0.01114	0.26899	0.14340	0.00000	0.00000	0.00000
#05 multilearn T	0.85712	0.00000	0.55735	0.00013	0.41274	0.02102	0.88889	0.74510	0.70588
#06 multilearn+forgets T	0.85493	0.00000	0.55678	0.00013	0.40899	0.02234	0.88889	0.74510	0.70588
#07 multigs+multilearn T	0.75735	0.00000	0.42352	0.00519	0.29474	0.10749	0.80000	0.70175	0.66667
#08 multigs+multilearn+forgets T	0.75674	0.00000	0.42217	0.00535	0.29469	0.10755	0.80000	0.70175	0.66667
#09 multiprior T	0.85945	0.00000	0.60892	0.00002	0.39342	0.02855	0.88889	0.74510	0.70588
#10 multiprior+forgets T	0.86665	0.00000	0.60710	0.00002	0.39300	0.02874	0.90566	0.76000	0.72000
#11 multigs TE	0.75714	0.00000	0.42340	0.00521	0.29603	0.10588	0.80000	0.70175	0.66667
#12 multigs+forgets TE	0.75689	0.00000	0.42372	0.00517	0.29814	0.10330	0.80000	0.70175	0.66667
#13 multilearn TE	0.85372	0.00000	0.55509	0.00014	0.40524	0.02372	0.88889	0.74510	0.70588
#14 multilearn+forgets TE	0.86909	0.00000	0.60680	0.00002	0.39485	0.02793	0.90566	0.76000	0.72000
#15 multigs+multilearn TE	0.75633	0.00000	0.42289	0.00527	0.29547	0.10657	0.80000	0.70175	0.66667
#16 multigs+multilearn+forgets TE	0.75811	0.00000	0.42294	0.00526	0.29586	0.10609	0.80000	0.70175	0.66667
#17 multiprior TE	0.85945	0.00000	0.60892	0.00002	0.39342	0.02855	0.88889	0.74510	0.70588
#18 multiprior+forgets TE	0.86665	0.00000	0.60710	0.00002	0.39300	0.02874	0.90566	0.76000	0.72000

Table C4. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 04.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85275	0.00000	0.40557	0.00942	0.58371	0.00071	0.32000	0.68182	0.65000
#02 vanilla+forgets	0.80468	0.00000	0.42580	0.00616	0.42379	0.01961	0.00000	0.00000	0.00000
#03 multigs T	0.84376	0.00000	0.39497	0.01166	0.50858	0.00411	0.28571	0.72340	0.65116
#04 multigs+forgets T	0.74160	0.00000	0.42075	0.00686	0.40303	0.02722	0.00000	0.00000	0.00000
#05 multilearn T	0.85285	0.00000	0.40567	0.00940	0.58337	0.00072	0.32000	0.68182	0.65000
#06 multilearn+forgets T	0.82648	0.00000	0.33089	0.03702	0.42012	0.02081	0.75000	0.22222	0.26087
#07 multigs+multilearn T	0.84357	0.00000	0.39320	0.01207	0.50970	0.00401	0.28571	0.72340	0.65116
#08 multigs+multilearn+forgets T	0.73942	0.00000	0.42085	0.00685	0.40138	0.02792	0.00000	0.00000	0.00000
#09 multiprior T	0.85609	0.00000	0.34503	0.02923	0.53141	0.00251	0.36364	0.58537	0.59459
#10 multiprior+forgets T	0.83535	0.00000	0.42823	0.00584	0.40949	0.02463	0.00000	0.00000	0.00000
#11 multigs TE	0.84523	0.00000	0.40160	0.01021	0.50381	0.00453	0.28571	0.72340	0.65116
#12 multigs+forgets TE	0.75434	0.00000	0.42177	0.00671	0.40861	0.02497	0.00000	0.00000	0.00000
#13 multilearn TE	0.85302	0.00000	0.40584	0.00937	0.58294	0.00072	0.32000	0.68182	0.65000
#14 multilearn+forgets TE	0.81549	0.00000	0.42762	0.00592	0.42071	0.02061	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.84554	0.00000	0.40131	0.01027	0.50416	0.00450	0.28571	0.72340	0.65116
#16 multigs+multilearn+forgets TE	0.79544	0.00000	0.42441	0.00634	0.41926	0.02110	0.00000	0.00000	0.00000
#17 multiprior TE	0.85609	0.00000	0.34503	0.02923	0.53141	0.00251	0.36364	0.58537	0.59459
#18 multiprior+forgets TE	0.83535	0.00000	0.42823	0.00584	0.40949	0.02463	0.00000	0.00000	0.00000

Table C5. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 05.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82586	0.00000	0.48685	0.00194	0.56088	0.00234	0.71111	0.64000	0.54545
#02 vanilla+forgets	0.81057	0.00000	0.47573	0.00254	0.55982	0.00239	0.71111	0.64000	0.54545
#03 multigs T	0.82153	0.00000	0.48053	0.00226	0.55903	0.00244	0.71111	0.64000	0.54545
#04 multigs+forgets T	0.78212	0.00000	0.46035	0.00363	0.56132	0.00232	0.68085	0.61538	0.52174
#05 multilearn T	0.82824	0.00000	0.47847	0.00238	0.56128	0.00232	0.71111	0.64000	0.54545
#06 multilearn+forgets T	0.82434	0.00000	0.48470	0.00204	0.56171	0.00230	0.71111	0.64000	0.54545
#07 multigs+multilearn T	0.81865	0.00000	0.47990	0.00230	0.55755	0.00252	0.71111	0.64000	0.54545
#08 multigs+multilearn+forgets T	0.77742	0.00000	0.45656	0.00396	0.55960	0.00241	0.68085	0.61538	0.52174
#09 multiprior T	0.84879	0.00000	0.45983	0.00368	0.62296	0.00052	0.72727	0.61224	0.55814
#10 multiprior+forgets T	0.84486	0.00000	0.45249	0.00434	0.62175	0.00054	0.72727	0.61224	0.55814
#11 multigs TE	0.82221	0.00000	0.47364	0.00267	0.55850	0.00246	0.71111	0.64000	0.54545
#12 multigs+forgets TE	0.79918	0.00000	0.46883	0.00299	0.55968	0.00240	0.71111	0.64000	0.54545
#13 multilearn TE	0.82638	0.00000	0.47837	0.00238	0.56150	0.00231	0.71111	0.64000	0.54545
#14 multilearn+forgets TE	0.81241	0.00000	0.47771	0.00242	0.56042	0.00236	0.71111	0.64000	0.54545
#15 multigs+multilearn TE	0.81922	0.00000	0.48193	0.00219	0.55903	0.00244	0.71111	0.64000	0.54545
#16 multigs+multilearn+forgets TE	0.79514	0.00000	0.46986	0.00292	0.56375	0.00220	0.71111	0.64000	0.54545
#17 multiprior TE	0.84879	0.00000	0.45983	0.00368	0.62296	0.00052	0.72727	0.61224	0.55814
#18 multiprior+forgets TE	0.84486	0.00000	0.45249	0.00434	0.62175	0.00054	0.72727	0.61224	0.55814

Table C6. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 06.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83820	0.00000	0.43007	0.00992	0.43075	0.02804	0.60000	0.46154	0.44444
#02 vanilla+forgets	0.84369	0.00000	0.42840	0.01024	0.42517	0.03037	0.60000	0.46154	0.44444
#03 multigs T	0.82343	0.00000	0.42917	0.01009	0.43599	0.02598	0.54545	0.51852	0.50000
#04 multigs+forgets T	0.82891	0.00000	0.42856	0.01021	0.43446	0.02657	0.60000	0.46154	0.44444
#05 multilearn T	0.83438	0.00000	0.43033	0.00987	0.43260	0.02729	0.60000	0.46154	0.44444
#06 multilearn+forgets T	0.84599	0.00000	0.42720	0.01048	0.42168	0.03190	0.60000	0.46154	0.44444
#07 multigs+multilearn T	0.82336	0.00000	0.42909	0.01011	0.43602	0.02597	0.54545	0.51852	0.50000
#08 multigs+multilearn+forgets T	0.82339	0.00000	0.42823	0.01028	0.43591	0.02601	0.54545	0.51852	0.50000
#09 multiprior T	0.88312	0.00000	0.44310	0.00768	0.43202	0.02753	0.66667	0.40000	0.38462
#10 multiprior+forgets T	0.89581	0.00000	0.43999	0.00817	0.41825	0.03347	0.66667	0.40000	0.38462
#11 multigs TE	0.82297	0.00000	0.42894	0.01014	0.43613	0.02592	0.54545	0.51852	0.50000
#12 multigs+forgets TE	0.82250	0.00000	0.42875	0.01017	0.43596	0.02599	0.54545	0.51852	0.50000
#13 multilearn TE	0.83438	0.00000	0.43033	0.00987	0.43260	0.02729	0.60000	0.46154	0.44444
#14 multilearn+forgets TE	0.84621	0.00000	0.42703	0.01052	0.42147	0.03200	0.60000	0.46154	0.44444
#15 multigs+multilearn TE	0.82248	0.00000	0.42923	0.01008	0.43609	0.02594	0.54545	0.51852	0.50000
#16 multigs+multilearn+forgets TE	0.82264	0.00000	0.42830	0.01026	0.43603	0.02596	0.54545	0.51852	0.50000
#17 multiprior TE	0.88312	0.00000	0.44310	0.00768	0.43202	0.02753	0.66667	0.40000	0.38462
#18 multiprior+forgets TE	0.89581	0.00000	0.43999	0.00817	0.41825	0.03347	0.66667	0.40000	0.38462

Table C7. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 07.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.88699	0.00000	0.55323	0.00000	0.40416	0.00102	0.92437	0.81132	0.76471
#02 vanilla+forgets	0.88074	0.00000	0.54408	0.00000	0.38865	0.00165	0.93333	0.80374	0.75728
#03 multigs T	0.50463	0.00000	0.31192	0.00387	0.18830	0.13942	0.83221	0.72059	0.66667
#04 multigs+forgets T	0.33004	0.00109	0.21465	0.04991	0.24312	0.05486	0.80519	0.69504	0.65693
#05 multilearn T	0.88397	0.00000	0.55340	0.00000	0.40077	0.00113	0.92437	0.81132	0.76471
#06 multilearn+forgets T	0.88109	0.00000	0.54541	0.00000	0.38956	0.00160	0.93333	0.80374	0.75728
#07 multigs+multilearn T	0.52591	0.00000	0.31167	0.00390	0.17877	0.16097	0.83784	0.72593	0.67176
#08 multigs+multilearn+forgets T	0.35213	0.00047	0.22604	0.03869	0.25235	0.04602	0.80519	0.69504	0.65693
#09 multiprior T	0.88537	0.00000	0.55143	0.00000	0.43695	0.00034	0.92437	0.81132	0.76471
#10 multiprior+forgets T	0.88935	0.00000	0.53818	0.00000	0.42371	0.00054	0.90435	0.78431	0.75510
#11 multigs TE	0.50492	0.00000	0.32171	0.00284	0.20023	0.11562	0.83221	0.72059	0.66667
#12 multigs+forgets TE	0.80429	0.00000	0.44943	0.00002	0.25584	0.04299	0.00000	0.00000	0.00000
#13 multilearn TE	0.88953	0.00000	0.56334	0.00000	0.40351	0.00104	0.93220	0.81905	0.77228
#14 multilearn+forgets TE	0.88578	0.00000	0.54808	0.00000	0.39229	0.00147	0.92437	0.81132	0.76471
#15 multigs+multilearn TE	0.50835	0.00000	0.32014	0.00299	0.19095	0.13385	0.83221	0.72059	0.66667
#16 multigs+multilearn+forgets TE	0.34625	0.00059	0.22762	0.03731	0.25783	0.04134	0.80519	0.69504	0.65693
#17 multiprior TE	0.88537	0.00000	0.55143	0.00000	0.43695	0.00034	0.92437	0.81132	0.76471
#18 multiprior+forgets TE	0.88935	0.00000	0.53818	0.00000	0.42371	0.00054	0.90435	0.78431	0.75510

Table C8. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 08.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82394	0.00000	0.56228	0.00000	0.42041	0.00172	0.51163	0.68571	0.65625
#02 vanilla+forgets	0.71100	0.00000	0.44387	0.00005	0.33791	0.01334	0.00000	0.00000	0.00000
#03 multigs T	0.78147	0.00000	0.50689	0.00000	0.35641	0.00881	0.44898	0.68421	0.62857
#04 multigs+forgets T	0.61965	0.00000	0.42918	0.00009	0.32232	0.01858	0.00000	0.00000	0.00000
#05 multilearn T	0.82376	0.00000	0.56250	0.00000	0.42090	0.00170	0.51163	0.68571	0.65625
#06 multilearn+forgets T	0.69630	0.00000	0.43282	0.00008	0.32889	0.01619	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.78152	0.00000	0.50692	0.00000	0.35648	0.00879	0.44898	0.68421	0.62857
#08 multigs+multilearn+forgets T	0.74793	0.00000	0.50095	0.00000	0.36781	0.00674	0.00000	0.00000	0.00000
#09 multiprior T	0.81941	0.00000	0.48781	0.00001	0.37037	0.00634	0.75862	0.50000	0.48000
#10 multiprior+forgets T	0.64662	0.00000	0.39152	0.00039	0.29770	0.03039	0.00000	0.00000	0.00000
#11 multigs TE	0.80383	0.00000	0.51291	0.00000	0.37851	0.00519	0.47826	0.65753	0.62687
#12 multigs+forgets TE	0.79852	0.00000	0.52461	0.00000	0.38064	0.00493	0.00000	0.00000	0.00000
#13 multilearn TE	0.82375	0.00000	0.56249	0.00000	0.42089	0.00170	0.51163	0.68571	0.65625
#14 multilearn+forgets TE	0.68214	0.00000	0.41682	0.00015	0.31220	0.02285	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.80378	0.00000	0.51289	0.00000	0.37845	0.00520	0.47826	0.65753	0.62687
#16 multigs+multilearn+forgets TE	0.58054	0.00000	0.36852	0.00090	0.27117	0.04953	0.00000	0.00000	0.00000
#17 multiprior TE	0.81941	0.00000	0.48781	0.00001	0.37037	0.00634	0.75862	0.50000	0.48000
#18 multiprior+forgets TE	0.64662	0.00000	0.39152	0.00039	0.29770	0.03039	0.00000	0.00000	0.00000

Table C9. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 09.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.77224	0.00000	0.39900	0.00011	0.37651	0.00236	0.52174	0.73043	0.67890
#02 vanilla+forgets	0.67883	0.00000	0.32727	0.00174	0.20537	0.10638	0.00000	0.00000	0.00000
#03 multigs T	0.71474	0.00000	0.38599	0.00019	0.47930	0.00007	0.50000	0.72269	0.70796
#04 multigs+forgets T	0.68175	0.00000	0.32224	0.00207	0.20468	0.10758	0.00000	0.00000	0.00000
#05 multilearn T	0.77259	0.00000	0.39905	0.00011	0.37646	0.00236	0.52174	0.73043	0.67890
#06 multilearn+forgets T	0.77350	0.00000	0.45059	0.00001	0.39311	0.00144	0.55172	0.76364	0.71154
#07 multigs+multilearn T	0.71561	0.00000	0.38762	0.00017	0.47930	0.00007	0.50000	0.72269	0.70796
#08 multigs+multilearn+forgets T	0.77319	0.00000	0.40572	0.00008	0.40813	0.00090	0.55814	0.75229	0.67961
#09 multiprior T	0.82637	0.00000	0.55133	0.00000	0.44646	0.00024	0.88000	0.60274	0.56716
#10 multiprior+forgets T	0.62435	0.00000	0.31054	0.00306	0.25027	0.04789	0.00000	0.00000	0.00000
#11 multigs TE	0.73487	0.00000	0.38445	0.00020	0.42484	0.00052	0.51064	0.71795	0.70270
#12 multigs+forgets TE	0.69888	0.00000	0.33310	0.00142	0.20770	0.10238	0.00000	0.00000	0.00000
#13 multilearn TE	0.77261	0.00000	0.39906	0.00011	0.37644	0.00236	0.52174	0.73043	0.67890
#14 multilearn+forgets TE	0.67143	0.00000	0.32607	0.00182	0.21049	0.09775	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.73489	0.00000	0.38447	0.00020	0.42482	0.00052	0.51064	0.71795	0.70270
#16 multigs+multilearn+forgets TE	0.59150	0.00000	0.31820	0.00237	0.26057	0.03915	0.00000	0.00000	0.00000
#17 multiprior TE	0.82637	0.00000	0.55133	0.00000	0.44646	0.00024	0.88000	0.60274	0.56716
#18 multiprior+forgets TE	0.62435	0.00000	0.31054	0.00306	0.25027	0.04789	0.00000	0.00000	0.00000

Table C10. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 10.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.80750	0.00000	0.33257	0.00313	0.37269	0.00707	0.94656	0.65421	0.59406
#02 vanilla+forgets	0.81494	0.00000	0.34244	0.00230	0.37137	0.00730	0.94574	0.66667	0.60606
#03 multigs T	0.60415	0.00000	0.27752	0.01454	0.28050	0.04618	0.87324	0.64407	0.57143
#04 multigs+forgets T	0.60415	0.00000	0.27752	0.01454	0.28050	0.04618	0.87324	0.64407	0.57143
#05 multilearn T	0.82091	0.00000	0.33095	0.00328	0.37580	0.00657	0.94574	0.66667	0.60606
#06 multilearn+forgets T	0.83136	0.00000	0.35527	0.00152	0.37557	0.00661	0.95312	0.67308	0.61224
#07 multigs+multilearn T	0.61318	0.00000	0.28816	0.01104	0.28364	0.04369	0.87324	0.64407	0.57143
#08 multigs+multilearn+forgets T	0.60435	0.00000	0.27757	0.01452	0.28340	0.04388	0.87324	0.64407	0.57143
#09 multiprior T	0.80274	0.00000	0.33158	0.00322	0.36662	0.00814	0.93846	0.66038	0.60000
#10 multiprior+forgets T	0.83060	0.00000	0.37263	0.00085	0.37335	0.00697	0.94309	0.70707	0.64516
#11 multigs TE	0.62228	0.00000	0.29285	0.00975	0.29598	0.03496	0.87324	0.64407	0.57143
#12 multigs+forgets TE	0.61930	0.00000	0.28180	0.01303	0.28480	0.04280	0.87324	0.64407	0.57143
#13 multilearn TE	0.81791	0.00000	0.32974	0.00341	0.37329	0.00698	0.94574	0.66667	0.60606
#14 multilearn+forgets TE	0.83952	0.00000	0.35432	0.00157	0.38207	0.00566	0.96063	0.67961	0.61856
#15 multigs+multilearn TE	0.62238	0.00000	0.29304	0.00970	0.29644	0.03466	0.87324	0.64407	0.57143
#16 multigs+multilearn+forgets TE	0.61404	0.00000	0.28379	0.01238	0.28688	0.04124	0.87324	0.64407	0.57143
#17 multiprior TE	0.80274	0.00000	0.33158	0.00322	0.36662	0.00814	0.93846	0.66038	0.60000
#18 multiprior+forgets TE	0.83060	0.00000	0.37263	0.00085	0.37335	0.00697	0.94309	0.70707	0.64516

Table C11. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 11.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.84176	0.00000	0.72746	0.00000	0.49948	0.00019	0.78788	0.79487	0.66667
#02 vanilla+forgets	0.81233	0.00000	0.67435	0.00000	0.59543	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.75935	0.00000	0.66232	0.00000	0.54745	0.00003	0.70270	0.76744	0.67500
#04 multigs+forgets T	0.73754	0.00000	0.61004	0.00000	0.48163	0.00035	0.67532	0.76404	0.67470
#05 multilearn T	0.84140	0.00000	0.72672	0.00000	0.49826	0.00020	0.78788	0.79487	0.66667
#06 multilearn+forgets T	0.81138	0.00000	0.73374	0.00000	0.50160	0.00018	0.76190	0.77333	0.63768
#07 multigs+multilearn T	0.75115	0.00000	0.63817	0.00000	0.54608	0.00003	0.69333	0.75862	0.66667
#08 multigs+multilearn+forgets T	0.74832	0.00000	0.64000	0.00000	0.54197	0.00004	0.68421	0.77273	0.68293
#09 multiprior T	0.86736	0.00000	0.76187	0.00000	0.61145	0.00000	0.83871	0.78378	0.67647
#10 multiprior+forgets T	0.80560	0.00000	0.66847	0.00000	0.58592	0.00001	0.00000	0.00000	0.00000
#11 multigs TE	0.75087	0.00000	0.63800	0.00000	0.54808	0.00003	0.69333	0.75862	0.66667
#12 multigs+forgets TE	0.73735	0.00000	0.61021	0.00000	0.47980	0.00037	0.67532	0.76404	0.67470
#13 multilearn TE	0.84154	0.00000	0.72696	0.00000	0.49862	0.00020	0.78788	0.79487	0.66667
#14 multilearn+forgets TE	0.82212	0.00000	0.67902	0.00000	0.52182	0.00009	0.66667	0.53333	0.48148
#15 multigs+multilearn TE	0.75002	0.00000	0.63807	0.00000	0.54852	0.00003	0.69333	0.75862	0.66667
#16 multigs+multilearn+forgets TE	0.73731	0.00000	0.61050	0.00000	0.47573	0.00042	0.67532	0.76404	0.67470
#17 multiprior TE	0.86736	0.00000	0.76187	0.00000	0.61145	0.00000	0.83871	0.78378	0.67647
#18 multiprior+forgets TE	0.80560	0.00000	0.66847	0.00000	0.58592	0.00001	0.00000	0.00000	0.00000

Table C12. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 12.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.81115	0.00000	0.49787	0.00002	0.51653	0.00020	0.83077	0.63889	0.63768
#02 vanilla+forgets	0.80721	0.00000	0.58396	0.00000	0.67171	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.68213	0.00000	0.48966	0.00003	0.43149	0.00246	0.63529	0.67391	0.65169
#04 multigs+forgets T	0.63242	0.00000	0.43037	0.00031	0.45239	0.00141	0.60674	0.66667	0.64516
#05 multilearn T	0.81065	0.00000	0.49759	0.00002	0.51487	0.00021	0.83077	0.63889	0.63768
#06 multilearn+forgets T	0.81104	0.00000	0.58345	0.00000	0.66689	0.00000	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.68160	0.00000	0.48647	0.00003	0.43058	0.00252	0.63529	0.67391	0.65169
#08 multigs+multilearn+forgets T	0.64684	0.00000	0.41497	0.00053	0.44940	0.00153	0.61364	0.65263	0.63043
#09 multiprior T	0.84803	0.00000	0.58805	0.00000	0.51075	0.00024	0.88462	0.67797	0.60714
#10 multiprior+forgets T	0.76766	0.00000	0.56652	0.00000	0.65257	0.00000	0.00000	0.00000	0.00000
#11 multigs TE	0.68372	0.00000	0.48742	0.00003	0.43694	0.00213	0.63529	0.67391	0.65169
#12 multigs+forgets TE	0.60286	0.00000	0.37310	0.00203	0.28909	0.04874	0.59341	0.65306	0.63158
#13 multilearn TE	0.81105	0.00000	0.49676	0.00002	0.51537	0.00021	0.83077	0.63889	0.63768
#14 multilearn+forgets TE	0.79304	0.00000	0.57669	0.00000	0.65658	0.00000	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.68125	0.00000	0.48001	0.00005	0.43281	0.00238	0.63529	0.67391	0.65169
#16 multigs+multilearn+forgets TE	0.64190	0.00000	0.41568	0.00052	0.44308	0.00181	0.64286	0.63736	0.63636
#17 multiprior TE	0.84803	0.00000	0.58805	0.00000	0.51075	0.00024	0.88462	0.67797	0.60714
#18 multiprior+forgets TE	0.76766	0.00000	0.56652	0.00000	0.65257	0.00000	0.00000	0.00000	0.00000

Table C13. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 13.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83308	0.00000	0.65053	0.00000	0.51297	0.00014	0.74074	0.77143	0.66667
#02 vanilla+forgets	0.78675	0.00000	0.57858	0.00000	0.55530	0.00003	0.00000	0.00000	0.00000
#03 multigs T	0.82983	0.00000	0.66493	0.00000	0.51754	0.00012	0.71429	0.77778	0.67647
#04 multigs+forgets T	0.78156	0.00000	0.57456	0.00000	0.55679	0.00003	0.00000	0.00000	0.00000
#05 multilearn T	0.83289	0.00000	0.64652	0.00000	0.51396	0.00013	0.74074	0.77143	0.66667
#06 multilearn+forgets T	0.78994	0.00000	0.58064	0.00000	0.55508	0.00003	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.82985	0.00000	0.66050	0.00000	0.51563	0.00013	0.72727	0.78873	0.68657
#08 multigs+multilearn+forgets T	0.78654	0.00000	0.57787	0.00000	0.55704	0.00003	0.00000	0.00000	0.00000
#09 multiprior T	0.84275	0.00000	0.64514	0.00000	0.49540	0.00025	0.74074	0.74286	0.66667
#10 multiprior+forgets T	0.73183	0.00000	0.53034	0.00000	0.54424	0.00004	0.00000	0.00000	0.00000
#11 multigs TE	0.76723	0.00000	0.62108	0.00000	0.49042	0.00030	0.61538	0.79012	0.67532
#12 multigs+forgets TE	0.67108	0.00000	0.45085	0.00013	0.50879	0.00016	0.00000	0.00000	0.00000
#13 multilearn TE	0.83274	0.00000	0.64627	0.00000	0.51417	0.00013	0.74074	0.77143	0.66667
#14 multilearn+forgets TE	0.79687	0.00000	0.58518	0.00000	0.55491	0.00003	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.76880	0.00000	0.62181	0.00000	0.48870	0.00032	0.61538	0.79012	0.67532
#16 multigs+multilearn+forgets TE	0.77544	0.00000	0.62272	0.00000	0.47011	0.00057	0.68966	0.78378	0.65714
#17 multiprior TE	0.84275	0.00000	0.64514	0.00000	0.49540	0.00025	0.74074	0.74286	0.66667
#18 multiprior+forgets TE	0.73183	0.00000	0.53034	0.00000	0.54424	0.00004	0.00000	0.00000	0.00000

Table C14. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 14.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83730	0.00000	0.54977	0.00000	0.43517	0.00159	0.84337	0.71429	0.61538
#02 vanilla+forgets	0.75985	0.00000	0.61641	0.00000	0.59979	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.68222	0.00000	0.38865	0.00126	0.28524	0.04466	0.72917	0.68041	0.61538
#04 multigs+forgets T	0.61189	0.00000	0.32753	0.00726	0.29591	0.03694	0.70000	0.67327	0.61053
#05 multilearn T	0.83686	0.00000	0.54295	0.00000	0.42815	0.00192	0.84337	0.71429	0.61538
#06 multilearn+forgets T	0.83959	0.00000	0.64124	0.00000	0.54053	0.00005	0.73684	0.65517	0.53846
#07 multigs+multilearn T	0.68238	0.00000	0.38873	0.00126	0.28541	0.04452	0.72917	0.68041	0.61538
#08 multigs+multilearn+forgets T	0.63364	0.00000	0.35895	0.00308	0.30506	0.03123	0.71429	0.66667	0.60215
#09 multiprior T	0.85986	0.00000	0.58466	0.00000	0.48237	0.00039	0.87500	0.71605	0.64000
#10 multiprior+forgets T	0.73304	0.00000	0.59632	0.00000	0.58680	0.00001	0.00000	0.00000	0.00000
#11 multigs TE	0.68221	0.00000	0.38868	0.00126	0.28525	0.04465	0.72917	0.68041	0.61538
#12 multigs+forgets TE	0.61393	0.00000	0.32969	0.00687	0.29826	0.03540	0.70000	0.67327	0.61053
#13 multilearn TE	0.83748	0.00000	0.54083	0.00000	0.42603	0.00204	0.84337	0.71429	0.61538
#14 multilearn+forgets TE	0.85365	0.00000	0.65566	0.00000	0.60926	0.00000	0.81967	0.64516	0.64286
#15 multigs+multilearn TE	0.68205	0.00000	0.38948	0.00123	0.28590	0.04414	0.72917	0.68041	0.61538
#16 multigs+multilearn+forgets TE	0.63347	0.00000	0.36119	0.00289	0.30958	0.02869	0.70707	0.68000	0.61702
#17 multiprior TE	0.85986	0.00000	0.58466	0.00000	0.48237	0.00039	0.87500	0.71605	0.64000
#18 multiprior+forgets TE	0.73304	0.00000	0.59632	0.00000	0.58680	0.00001	0.00000	0.00000	0.00000

Table C15. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 15.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.86284	0.00000	0.34548	0.00276	0.23497	0.09034	0.90722	0.64368	0.60465
#02 vanilla+forgets	0.84526	0.00000	0.34069	0.00318	0.17176	0.21878	0.89796	0.63636	0.59770
#03 multigs T	0.74007	0.00000	0.33074	0.00426	0.12669	0.36602	0.82243	0.59794	0.60417
#04 multigs+forgets T	0.71233	0.00000	0.30277	0.00922	0.11434	0.41494	0.81481	0.59184	0.59794
#05 multilearn T	0.88474	0.00000	0.34323	0.00295	0.21985	0.11369	0.91667	0.65116	0.61176
#06 multilearn+forgets T	0.86165	0.00000	0.34148	0.00311	0.22478	0.10563	0.90722	0.64368	0.60465
#07 multigs+multilearn T	0.74024	0.00000	0.33054	0.00429	0.12665	0.36617	0.82243	0.59794	0.60417
#08 multigs+multilearn+forgets T	0.73492	0.00000	0.33432	0.00384	0.12096	0.38826	0.82243	0.59794	0.60417
#09 multiprior T	0.85650	0.00000	0.33593	0.00367	0.22344	0.10777	0.90722	0.64368	0.60465
#10 multiprior+forgets T	0.82335	0.00000	0.30745	0.00815	0.18197	0.19221	0.88000	0.62222	0.60674
#11 multigs TE	0.73906	0.00000	0.32768	0.00466	0.13274	0.34338	0.83019	0.60417	0.58947
#12 multigs+forgets TE	0.73867	0.00000	0.33057	0.00429	0.12563	0.37006	0.82243	0.59794	0.60417
#13 multilearn TE	0.88328	0.00000	0.34273	0.00300	0.22208	0.10999	0.91667	0.65116	0.61176
#14 multilearn+forgets TE	0.88602	0.00000	0.34725	0.00261	0.20454	0.14179	0.93617	0.64286	0.57831
#15 multigs+multilearn TE	0.73866	0.00000	0.32805	0.00461	0.13134	0.34853	0.83019	0.60417	0.58947
#16 multigs+multilearn+forgets TE	0.74585	0.00000	0.33347	0.00394	0.12234	0.38283	0.82243	0.59794	0.60417
#17 multiprior TE	0.85650	0.00000	0.33593	0.00367	0.22344	0.10777	0.90722	0.64368	0.60465
#18 multiprior+forgets TE	0.82335	0.00000	0.30745	0.00815	0.18197	0.19221	0.88000	0.62222	0.60674

Table C16. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 16.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83454	0.00000	0.46822	0.00003	0.38069	0.00537	0.73333	0.55556	0.44444
#02 vanilla+forgets	0.89157	0.00000	0.43886	0.00010	0.44387	0.00098	0.00000	0.00000	0.00000
#03 multigs T	0.69778	0.00000	0.46021	0.00004	0.32064	0.02048	0.50000	0.67647	0.61765
#04 multigs+forgets T	0.84026	0.00000	0.38749	0.00071	0.39019	0.00424	0.00000	0.00000	0.00000
#05 multilearn T	0.83475	0.00000	0.46829	0.00003	0.38088	0.00534	0.73333	0.55556	0.44444
#06 multilearn+forgets T	0.88114	0.00000	0.42556	0.00017	0.43435	0.00129	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.69813	0.00000	0.46042	0.00004	0.32006	0.02072	0.50000	0.67647	0.61765
#08 multigs+multilearn+forgets T	0.84243	0.00000	0.38132	0.00087	0.39291	0.00396	0.00000	0.00000	0.00000
#09 multiprior T	0.87005	0.00000	0.43159	0.00014	0.33593	0.01490	0.80000	0.44898	0.36735
#10 multiprior+forgets T	0.87583	0.00000	0.43913	0.00010	0.43737	0.00119	0.00000	0.00000	0.00000
#11 multigs TE	0.69234	0.00000	0.44284	0.00009	0.34894	0.01124	0.48889	0.66667	0.63768
#12 multigs+forgets TE	0.81673	0.00000	0.36241	0.00163	0.37060	0.00684	0.00000	0.00000	0.00000
#13 multilearn TE	0.83471	0.00000	0.46828	0.00003	0.38085	0.00535	0.73333	0.55556	0.44444
#14 multilearn+forgets TE	0.87405	0.00000	0.42675	0.00017	0.43206	0.00138	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.69218	0.00000	0.44315	0.00009	0.34916	0.01118	0.48889	0.66667	0.63768
#16 multigs+multilearn+forgets TE	0.82434	0.00000	0.37287	0.00116	0.37324	0.00642	0.00000	0.00000	0.00000
#17 multiprior TE	0.87005	0.00000	0.43159	0.00014	0.33593	0.01490	0.80000	0.44898	0.36735
#18 multiprior+forgets TE	0.87583	0.00000	0.43913	0.00010	0.43737	0.00119	0.00000	0.00000	0.00000

Table C17. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 17.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85291	0.00000	0.49887	0.00002	0.23296	0.10721	0.93878	0.79121	0.69767
#02 vanilla+forgets	0.81972	0.00000	0.47979	0.00005	0.15529	0.28669	0.94000	0.77419	0.68182
#03 multigs T	0.57450	0.00000	0.43072	0.00031	0.29485	0.03972	0.83186	0.73585	0.69307
#04 multigs+forgets T	0.47119	0.00002	0.38602	0.00137	0.27737	0.05367	0.81739	0.72222	0.67961
#05 multilearn T	0.85282	0.00000	0.49901	0.00002	0.23300	0.10714	0.93878	0.79121	0.69767
#06 multilearn+forgets T	0.84299	0.00000	0.46958	0.00007	0.21113	0.14534	0.91667	0.76404	0.69048
#07 multigs+multilearn T	0.57567	0.00000	0.42299	0.00040	0.29808	0.03751	0.83186	0.73585	0.69307
#08 multigs+multilearn+forgets T	0.20266	0.08333	0.18414	0.13886	0.17080	0.24064	0.78333	0.70796	0.64815
#09 multiprior T	0.88132	0.00000	0.50641	0.00001	0.14154	0.33198	0.91304	0.75294	0.67500
#10 multiprior+forgets T	0.88216	0.00000	0.51502	0.00001	0.14676	0.31428	0.92473	0.74419	0.66667
#11 multigs TE	0.67421	0.00000	0.42135	0.00043	0.11729	0.42220	0.86239	0.74510	0.65979
#12 multigs+forgets TE	0.62369	0.00000	0.37233	0.00208	0.17989	0.21616	0.84685	0.73077	0.66667
#13 multilearn TE	0.85292	0.00000	0.49907	0.00002	0.23301	0.10712	0.93878	0.79121	0.69767
#14 multilearn+forgets TE	0.84499	0.00000	0.47657	0.00005	0.21107	0.14544	0.91667	0.76404	0.69048
#15 multigs+multilearn TE	0.67484	0.00000	0.42036	0.00044	0.11826	0.41836	0.86239	0.74510	0.65979
#16 multigs+multilearn+forgets TE	0.72447	0.00000	0.48975	0.00003	0.26886	0.06176	0.87850	0.76000	0.67368
#17 multiprior TE	0.88132	0.00000	0.50641	0.00001	0.14154	0.33198	0.91304	0.75294	0.67500
#18 multiprior+forgets TE	0.88216	0.00000	0.51502	0.00001	0.14676	0.31428	0.92473	0.74419	0.66667

Table C18. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 19.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.81196	0.00000	0.47891	0.00004	0.26727	0.05543	0.68000	0.68571	0.68571
#02 vanilla+forgets	0.73069	0.00000	0.34970	0.00346	0.35571	0.00965	0.00000	0.00000	0.00000
#03 multigs T	0.71581	0.00000	0.49210	0.00002	0.30343	0.02877	0.57627	0.73418	0.73418
#04 multigs+forgets T	0.74277	0.00000	0.35436	0.00303	0.36766	0.00733	0.00000	0.00000	0.00000
#05 multilearn T	0.81137	0.00000	0.47922	0.00004	0.26710	0.05560	0.68000	0.68571	0.68571
#06 multilearn+forgets T	0.75805	0.00000	0.36785	0.00203	0.37120	0.00674	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.71689	0.00000	0.49255	0.00002	0.30311	0.02894	0.57627	0.73418	0.73418
#08 multigs+multilearn+forgets T	0.79367	0.00000	0.39975	0.00073	0.40563	0.00285	0.00000	0.00000	0.00000
#09 multiprior T	0.83446	0.00000	0.44320	0.00015	0.22936	0.10192	0.75556	0.64615	0.61538
#10 multiprior+forgets T	0.70156	0.00000	0.33155	0.00575	0.33943	0.01383	0.00000	0.00000	0.00000
#11 multigs TE	0.71601	0.00000	0.49236	0.00002	0.30361	0.02866	0.57627	0.73418	0.73418
#12 multigs+forgets TE	0.80064	0.00000	0.40649	0.00058	0.40925	0.00259	0.00000	0.00000	0.00000
#13 multilearn TE	0.81140	0.00000	0.47920	0.00004	0.26710	0.05559	0.68000	0.68571	0.68571
#14 multilearn+forgets TE	0.73523	0.00000	0.35630	0.00286	0.36010	0.00874	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.71565	0.00000	0.49200	0.00002	0.30343	0.02876	0.57627	0.73418	0.73418
#16 multigs+multilearn+forgets TE	0.75755	0.00000	0.53362	0.00000	0.32702	0.01797	0.68000	0.71429	0.71429
#17 multiprior TE	0.83446	0.00000	0.44320	0.00015	0.22936	0.10192	0.75556	0.64615	0.61538
#18 multiprior+forgets TE	0.70156	0.00000	0.33155	0.00575	0.33943	0.01383	0.00000	0.00000	0.00000

Table C19. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 21.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.89797	0.00000	0.32172	0.00622	0.33822	0.00941	0.94949	0.73913	0.70588
#02 vanilla+forgets	0.92404	0.00000	0.30909	0.00872	0.30052	0.02190	0.95918	0.72527	0.69048
#03 multigs T	0.83934	0.00000	0.30369	0.01004	0.32290	0.01343	0.91262	0.72917	0.67416
#04 multigs+forgets T	0.81846	0.00000	0.30116	0.01071	0.35479	0.00628	0.90385	0.72165	0.68889
#05 multilearn T	0.89723	0.00000	0.32031	0.00646	0.33845	0.00936	0.94949	0.73913	0.70588
#06 multilearn+forgets T	0.91970	0.00000	0.29810	0.01157	0.29501	0.02457	0.95918	0.72527	0.69048
#07 multigs+multilearn T	0.83934	0.00000	0.30368	0.01004	0.32290	0.01343	0.91262	0.72917	0.67416
#08 multigs+multilearn+forgets T	0.82317	0.00000	0.30931	0.00867	0.36617	0.00470	0.90385	0.72165	0.68889
#09 multiprior T	0.94164	0.00000	0.30877	0.00880	0.24100	0.06839	0.98947	0.72727	0.66667
#10 multiprior+forgets T	0.94191	0.00000	0.33991	0.00373	0.22893	0.08389	1.00000	0.73563	0.65000
#11 multigs TE	0.83937	0.00000	0.30367	0.01004	0.32288	0.01343	0.91262	0.72917	0.67416
#12 multigs+forgets TE	0.81851	0.00000	0.30121	0.01069	0.35479	0.00628	0.90385	0.72165	0.68889
#13 multilearn TE	0.89705	0.00000	0.32000	0.00652	0.33849	0.00935	0.94949	0.73913	0.70588
#14 multilearn+forgets TE	0.89114	0.00000	0.31382	0.00770	0.34928	0.00720	0.95833	0.71910	0.68293
#15 multigs+multilearn TE	0.83937	0.00000	0.30367	0.01004	0.32288	0.01343	0.91262	0.72917	0.67416
#16 multigs+multilearn+forgets TE	0.86670	0.00000	0.30425	0.00989	0.37280	0.00395	0.92000	0.73118	0.69767
#17 multiprior TE	0.94164	0.00000	0.30877	0.00880	0.24100	0.06839	0.98947	0.72727	0.66667
#18 multiprior+forgets TE	0.94191	0.00000	0.33991	0.00373	0.22893	0.08389	1.00000	0.73563	0.65000

Table C20. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 23.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85185	0.00000	0.32251	0.01442	0.48306	0.00038	0.92500	0.68421	0.71233
#02 vanilla+forgets	0.85660	0.00000	0.33338	0.01127	0.48452	0.00036	0.93671	0.69333	0.72222
#03 multigs T	0.80687	0.00000	0.28546	0.03137	0.44014	0.00138	0.90244	0.69231	0.69333
#04 multigs+forgets T	0.80741	0.00000	0.28412	0.03220	0.44152	0.00133	0.90244	0.69231	0.69333
#05 multilearn T	0.85282	0.00000	0.32552	0.01348	0.48134	0.00040	0.92500	0.68421	0.71233
#06 multilearn+forgets T	0.85795	0.00000	0.33989	0.00969	0.47923	0.00043	0.93671	0.69333	0.72222
#07 multigs+multilearn T	0.80920	0.00000	0.29096	0.02811	0.43625	0.00154	0.90244	0.69231	0.69333
#08 multigs+multilearn+forgets T	0.81169	0.00000	0.29301	0.02697	0.43715	0.00150	0.90244	0.69231	0.69333
#09 multiprior T	0.89045	0.00000	0.27362	0.03945	0.48258	0.00039	0.96104	0.68493	0.74286
#10 multiprior+forgets T	0.90137	0.00000	0.30770	0.01989	0.48948	0.00031	0.96104	0.68493	0.74286
#11 multigs TE	0.80466	0.00000	0.28123	0.03407	0.44258	0.00129	0.90244	0.69231	0.69333
#12 multigs+forgets TE	0.80739	0.00000	0.28401	0.03227	0.44153	0.00133	0.90244	0.69231	0.69333
#13 multilearn TE	0.85188	0.00000	0.32259	0.01439	0.48302	0.00038	0.92500	0.68421	0.71233
#14 multilearn+forgets TE	0.85800	0.00000	0.33821	0.01008	0.48376	0.00037	0.93671	0.69333	0.72222
#15 multigs+multilearn TE	0.80587	0.00000	0.28315	0.03282	0.44140	0.00133	0.90244	0.69231	0.69333
#16 multigs+multilearn+forgets TE	0.80987	0.00000	0.28653	0.03071	0.44436	0.00123	0.90244	0.69231	0.69333
#17 multiprior TE	0.89045	0.00000	0.27362	0.03945	0.48258	0.00039	0.96104	0.68493	0.74286
#18 multiprior+forgets TE	0.90137	0.00000	0.30770	0.01989	0.48948	0.00031	0.96104	0.68493	0.74286

Table C21. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (testing subset) – DKC 25.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85347	0.00000	0.45077	0.00004	0.34409	0.00662	0.87059	0.74468	0.66667
#02 vanilla+forgets	0.84511	0.00000	0.45881	0.00003	0.33654	0.00800	0.88095	0.77419	0.65169
#03 multigs T	0.71045	0.00000	0.36705	0.00111	0.17728	0.17168	0.70476	0.73684	0.67273
#04 multigs+forgets T	0.80245	0.00000	0.47074	0.00002	0.34109	0.00714	0.00000	0.00000	0.00000
#05 multilearn T	0.84984	0.00000	0.45142	0.00004	0.35440	0.00507	0.87059	0.74468	0.66667
#06 multilearn+forgets T	0.85486	0.00000	0.47126	0.00002	0.34389	0.00666	0.91358	0.77778	0.65116
#07 multigs+multilearn T	0.71222	0.00000	0.36660	0.00113	0.17932	0.16672	0.70476	0.73684	0.67273
#08 multigs+multilearn+forgets T	0.80070	0.00000	0.47398	0.00002	0.33215	0.00892	0.00000	0.00000	0.00000
#09 multiprior T	0.84995	0.00000	0.45679	0.00003	0.33376	0.00857	0.89157	0.78261	0.65909
#10 multiprior+forgets T	0.78040	0.00000	0.47518	0.00001	0.36528	0.00380	0.00000	0.00000	0.00000
#11 multigs TE	0.70926	0.00000	0.36692	0.00111	0.17593	0.17502	0.70476	0.73684	0.67273
#12 multigs+forgets TE	0.71454	0.00000	0.37275	0.00091	0.17904	0.16740	0.70476	0.73684	0.67273
#13 multilearn TE	0.84940	0.00000	0.45116	0.00004	0.35405	0.00512	0.87059	0.74468	0.66667
#14 multilearn+forgets TE	0.85325	0.00000	0.46848	0.00002	0.34360	0.00670	0.91358	0.77778	0.65116
#15 multigs+multilearn TE	0.71261	0.00000	0.36690	0.00111	0.17977	0.16565	0.70476	0.73684	0.67273
#16 multigs+multilearn+forgets TE	0.78464	0.00000	0.47288	0.00002	0.32323	0.01106	0.00000	0.00000	0.00000
#17 multiprior TE	0.84995	0.00000	0.45679	0.00003	0.33376	0.00857	0.89157	0.78261	0.65909
#18 multiprior+forgets TE	0.78040	0.00000	0.47518	0.00001	0.36528	0.00380	0.00000	0.00000	0.00000

Table C22. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 01.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.86867	0.00000	0.36818	0.00072	0.41125	0.00081	0.95935	0.72072	0.69811
#02 vanilla+forgets	0.86164	0.00000	0.31124	0.00468	0.34941	0.00500	0.97521	0.69725	0.67308
#03 multigs T	0.56676	0.00000	0.21991	0.04853	0.17954	0.15914	0.79470	0.64748	0.59701
#04 multigs+forgets T	0.47617	0.00000	0.15488	0.16739	0.12817	0.31678	0.77419	0.64336	0.59420
#05 multilearn T	0.87706	0.00000	0.34672	0.00152	0.36052	0.00370	0.95798	0.71028	0.68627
#06 multilearn+forgets T	0.88323	0.00000	0.35411	0.00118	0.36555	0.00322	0.92174	0.69903	0.69388
#07 multigs+multilearn T	0.56771	0.00000	0.21987	0.04858	0.17940	0.15947	0.79470	0.64748	0.59701
#08 multigs+multilearn+forgets T	0.53085	0.00000	0.23218	0.03700	0.11066	0.38793	0.78947	0.64286	0.59259
#09 multiprior T	0.87551	0.00000	0.27082	0.01447	0.32238	0.00997	0.94017	0.66667	0.64000
#10 multiprior+forgets T	0.80320	0.00000	0.35196	0.00127	0.37421	0.00252	0.00000	0.00000	0.00000
#11 multigs TE	0.56251	0.00000	0.22087	0.04754	0.17973	0.15870	0.79470	0.64748	0.59701
#12 multigs+forgets TE	0.47397	0.00000	0.15423	0.16921	0.12750	0.31935	0.77419	0.64336	0.59420
#13 multilearn TE	0.87289	0.00000	0.34290	0.00173	0.37449	0.00250	0.95868	0.71560	0.69231
#14 multilearn+forgets TE	0.87411	0.00000	0.32616	0.00296	0.35432	0.00438	0.96667	0.70370	0.67961
#15 multigs+multilearn TE	0.56391	0.00000	0.21876	0.04976	0.17842	0.16181	0.79470	0.64748	0.59701
#16 multigs+multilearn+forgets TE	0.52919	0.00000	0.19742	0.07730	0.11737	0.35962	0.79470	0.64748	0.59701
#17 multiprior TE	0.87551	0.00000	0.27082	0.01447	0.32238	0.00997	0.94017	0.66667	0.64000
#18 multiprior+forgets TE	0.80320	0.00000	0.35196	0.00127	0.37421	0.00252	0.00000	0.00000	0.00000

Table C23. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 02.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.77419	0.00000	0.47243	0.00004	0.30477	0.02504	0.55000	0.75248	0.68750
#02 vanilla+forgets	0.69931	0.00000	0.41687	0.00037	0.33239	0.01406	0.00000	0.00000	0.00000
#03 multigs T	0.66105	0.00000	0.38998	0.00092	0.36229	0.00710	0.48352	0.71429	0.67290
#04 multigs+forgets T	0.62540	0.00000	0.36058	0.00234	0.37298	0.00547	0.00000	0.00000	0.00000
#05 multilearn T	0.77931	0.00000	0.48886	0.00002	0.31464	0.02049	0.55696	0.76000	0.69474
#06 multilearn+forgets T	0.70690	0.00000	0.42291	0.00029	0.32535	0.01637	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.66091	0.00000	0.38836	0.00098	0.36224	0.00711	0.48352	0.71429	0.67290
#08 multigs+multilearn+forgets T	0.62507	0.00000	0.36529	0.00203	0.36533	0.00660	0.00000	0.00000	0.00000
#09 multiprior T	0.85018	0.00000	0.55029	0.00000	0.40788	0.00220	0.84000	0.64789	0.60606
#10 multiprior+forgets T	0.60697	0.00000	0.35830	0.00250	0.27673	0.04280	0.00000	0.00000	0.00000
#11 multigs TE	0.66611	0.00000	0.39710	0.00073	0.35336	0.00877	0.48889	0.72072	0.66038
#12 multigs+forgets TE	0.67422	0.00000	0.39852	0.00069	0.35913	0.00766	0.00000	0.00000	0.00000
#13 multilearn TE	0.77415	0.00000	0.47220	0.00004	0.30325	0.02581	0.55000	0.75248	0.68750
#14 multilearn+forgets TE	0.69139	0.00000	0.41191	0.00044	0.33170	0.01427	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.66708	0.00000	0.39864	0.00069	0.35623	0.00820	0.48889	0.72072	0.66038
#16 multigs+multilearn+forgets TE	0.63534	0.00000	0.37343	0.00158	0.35572	0.00830	0.00000	0.00000	0.00000
#17 multiprior TE	0.85018	0.00000	0.55029	0.00000	0.40788	0.00220	0.84000	0.64789	0.60606
#18 multiprior+forgets TE	0.60697	0.00000	0.35830	0.00250	0.27673	0.04280	0.00000	0.00000	0.00000

Table C24. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 03.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83073	0.00000	0.43860	0.00046	0.32413	0.02798	0.88889	0.71429	0.64198
#02 vanilla+forgets	0.82830	0.00000	0.43497	0.00051	0.31464	0.03320	0.88889	0.71429	0.64198
#03 multigs T	0.75143	0.00000	0.37264	0.00337	0.29800	0.04428	0.83333	0.71111	0.64368
#04 multigs+forgets T	0.83305	0.00000	0.40443	0.00135	0.27148	0.06799	0.00000	0.00000	0.00000
#05 multilearn T	0.83763	0.00000	0.44092	0.00042	0.33219	0.02410	0.89888	0.72289	0.65000
#06 multilearn+forgets T	0.83613	0.00000	0.44249	0.00040	0.33035	0.02495	0.89888	0.72289	0.65000
#07 multigs+multilearn T	0.75189	0.00000	0.37270	0.00336	0.29666	0.04528	0.83333	0.71111	0.64368
#08 multigs+multilearn+forgets T	0.75188	0.00000	0.37210	0.00342	0.29575	0.04598	0.83333	0.71111	0.64368
#09 multiprior T	0.83884	0.00000	0.48165	0.00010	0.31526	0.03283	0.89888	0.72289	0.65000
#10 multiprior+forgets T	0.85275	0.00000	0.49554	0.00006	0.33709	0.02197	0.91954	0.74074	0.66667
#11 multigs TE	0.75172	0.00000	0.37191	0.00343	0.29679	0.04519	0.83333	0.71111	0.64368
#12 multigs+forgets TE	0.75198	0.00000	0.37182	0.00344	0.29501	0.04655	0.83333	0.71111	0.64368
#13 multilearn TE	0.83503	0.00000	0.43993	0.00044	0.32740	0.02635	0.89888	0.72289	0.65000
#14 multilearn+forgets TE	0.85389	0.00000	0.49565	0.00006	0.33908	0.02115	0.91954	0.74074	0.66667
#15 multigs+multilearn TE	0.75113	0.00000	0.37161	0.00346	0.29592	0.04585	0.83333	0.71111	0.64368
#16 multigs+multilearn+forgets TE	0.75231	0.00000	0.37100	0.00352	0.29438	0.04704	0.83333	0.71111	0.64368
#17 multiprior TE	0.83884	0.00000	0.48165	0.00010	0.31526	0.03283	0.89888	0.72289	0.65000
#18 multiprior+forgets TE	0.85275	0.00000	0.49554	0.00006	0.33709	0.02197	0.91954	0.74074	0.66667

Table C25. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 04.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85245	0.00000	0.45413	0.00039	0.42902	0.00366	0.24242	0.68852	0.65517
#02 vanilla+forgets	0.71523	0.00000	0.36691	0.00499	0.29134	0.05501	0.00000	0.00000	0.00000
#03 multigs T	0.82834	0.00000	0.48324	0.00014	0.40914	0.00582	0.20513	0.74627	0.68750
#04 multigs+forgets T	0.65756	0.00000	0.34509	0.00857	0.28300	0.06268	0.00000	0.00000	0.00000
#05 multilearn T	0.85261	0.00000	0.45445	0.00038	0.42926	0.00364	0.24242	0.68852	0.65517
#06 multilearn+forgets T	0.75438	0.00000	0.31599	0.01664	0.32508	0.03131	0.66667	0.21622	0.23529
#07 multigs+multilearn T	0.82772	0.00000	0.48112	0.00015	0.40990	0.00572	0.20513	0.74627	0.68750
#08 multigs+multilearn+forgets T	0.65526	0.00000	0.34458	0.00867	0.28157	0.06408	0.00000	0.00000	0.00000
#09 multiprior T	0.85427	0.00000	0.41969	0.00115	0.42020	0.00451	0.25806	0.64407	0.64286
#10 multiprior+forgets T	0.74060	0.00000	0.36771	0.00489	0.27493	0.07090	0.00000	0.00000	0.00000
#11 multigs TE	0.83365	0.00000	0.48751	0.00012	0.40497	0.00639	0.20513	0.74627	0.68750
#12 multigs+forgets TE	0.66963	0.00000	0.35000	0.00761	0.28560	0.06021	0.00000	0.00000	0.00000
#13 multilearn TE	0.85283	0.00000	0.45489	0.00038	0.42961	0.00361	0.24242	0.68852	0.65517
#14 multilearn+forgets TE	0.72445	0.00000	0.37170	0.00442	0.28860	0.05745	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.83389	0.00000	0.48698	0.00012	0.40495	0.00640	0.20513	0.74627	0.68750
#16 multigs+multilearn+forgets TE	0.70767	0.00000	0.36615	0.00509	0.29070	0.05558	0.00000	0.00000	0.00000
#17 multiprior TE	0.85427	0.00000	0.41969	0.00115	0.42020	0.00451	0.25806	0.64407	0.64286
#18 multiprior+forgets TE	0.74060	0.00000	0.36771	0.00489	0.27493	0.07090	0.00000	0.00000	0.00000

Table C26. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 05.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.80015	0.00000	0.55509	0.00001	0.48156	0.00081	0.68750	0.72000	0.67606
#02 vanilla+forgets	0.78836	0.00000	0.54658	0.00001	0.47925	0.00087	0.68750	0.72000	0.67606
#03 multigs T	0.79640	0.00000	0.55147	0.00001	0.47982	0.00085	0.68750	0.72000	0.67606
#04 multigs+forgets T	0.73806	0.00000	0.52834	0.00002	0.53807	0.00014	0.64706	0.70886	0.66667
#05 multilearn T	0.82919	0.00000	0.56670	0.00000	0.53766	0.00014	0.70968	0.73973	0.66667
#06 multilearn+forgets T	0.79844	0.00000	0.55241	0.00001	0.47995	0.00085	0.68750	0.72000	0.67606
#07 multigs+multilearn T	0.79412	0.00000	0.55004	0.00001	0.47938	0.00086	0.68750	0.72000	0.67606
#08 multigs+multilearn+forgets T	0.73251	0.00000	0.52363	0.00002	0.53546	0.00015	0.64706	0.70886	0.66667
#09 multiprior T	0.83548	0.00000	0.54622	0.00001	0.52726	0.00020	0.72131	0.69444	0.64706
#10 multiprior+forgets T	0.83346	0.00000	0.53760	0.00001	0.53232	0.00017	0.72131	0.69444	0.64706
#11 multigs TE	0.81031	0.00000	0.55466	0.00001	0.47306	0.00103	0.69841	0.72973	0.65714
#12 multigs+forgets TE	0.78079	0.00000	0.54506	0.00001	0.48423	0.00075	0.68750	0.72000	0.67606
#13 multilearn TE	0.81420	0.00000	0.55788	0.00000	0.47236	0.00105	0.69841	0.72973	0.65714
#14 multilearn+forgets TE	0.78964	0.00000	0.54804	0.00001	0.47949	0.00086	0.68750	0.72000	0.67606
#15 multigs+multilearn TE	0.79454	0.00000	0.55222	0.00001	0.48070	0.00083	0.68750	0.72000	0.67606
#16 multigs+multilearn+forgets TE	0.77795	0.00000	0.54702	0.00001	0.48886	0.00066	0.68750	0.72000	0.67606
#17 multiprior TE	0.83548	0.00000	0.54622	0.00001	0.52726	0.00020	0.72131	0.69444	0.64706
#18 multiprior+forgets TE	0.83346	0.00000	0.53760	0.00001	0.53232	0.00017	0.72131	0.69444	0.64706

Table C27. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 06.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82755	0.00000	0.30937	0.02419	0.36462	0.02071	0.55556	0.41860	0.45000
#02 vanilla+forgets	0.83031	0.00000	0.30726	0.02522	0.35953	0.02270	0.55556	0.41860	0.45000
#03 multigs T	0.82158	0.00000	0.31472	0.02172	0.37241	0.01796	0.52632	0.45455	0.48780
#04 multigs+forgets T	0.82460	0.00000	0.31373	0.02216	0.37036	0.01865	0.55556	0.41860	0.45000
#05 multilearn T	0.82566	0.00000	0.31048	0.02366	0.36679	0.01991	0.55556	0.41860	0.45000
#06 multilearn+forgets T	0.83129	0.00000	0.30576	0.02598	0.35638	0.02400	0.55556	0.41860	0.45000
#07 multigs+multilearn T	0.82152	0.00000	0.31464	0.02176	0.37243	0.01795	0.52632	0.45455	0.48780
#08 multigs+multilearn+forgets T	0.82268	0.00000	0.31656	0.02092	0.37478	0.01719	0.52632	0.45455	0.48780
#09 multiprior T	0.86217	0.00000	0.31027	0.02375	0.35281	0.02555	0.62500	0.39024	0.42105
#10 multiprior+forgets T	0.86969	0.00000	0.30570	0.02601	0.34023	0.03171	0.62500	0.39024	0.42105
#11 multigs TE	0.82120	0.00000	0.31436	0.02188	0.37256	0.01791	0.52632	0.45455	0.48780
#12 multigs+forgets TE	0.82184	0.00000	0.31644	0.02097	0.37460	0.01724	0.52632	0.45455	0.48780
#13 multilearn TE	0.82566	0.00000	0.31048	0.02366	0.36679	0.01991	0.55556	0.41860	0.45000
#14 multilearn+forgets TE	0.83140	0.00000	0.30560	0.02606	0.35617	0.02409	0.55556	0.41860	0.45000
#15 multigs+multilearn TE	0.82080	0.00000	0.31446	0.02184	0.37255	0.01791	0.52632	0.45455	0.48780
#16 multigs+multilearn+forgets TE	0.82190	0.00000	0.31606	0.02114	0.37463	0.01723	0.52632	0.45455	0.48780
#17 multiprior TE	0.86217	0.00000	0.31027	0.02375	0.35281	0.02555	0.62500	0.39024	0.42105
#18 multiprior+forgets TE	0.86969	0.00000	0.30570	0.02601	0.34023	0.03171	0.62500	0.39024	0.42105

Table C28. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 07.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.88096	0.00000	0.55066	0.00000	0.46115	0.00001	0.93642	0.78667	0.73239
#02 vanilla+forgets	0.88682	0.00000	0.55656	0.00000	0.46468	0.00001	0.94186	0.77852	0.73759
#03 multigs T	0.60113	0.00000	0.36824	0.00004	0.24742	0.02086	0.84762	0.70588	0.63687
#04 multigs+forgets T	0.45207	0.00000	0.28506	0.00168	0.25546	0.01694	0.81651	0.67692	0.62032
#05 multilearn T	0.88720	0.00000	0.56192	0.00000	0.47055	0.00000	0.93567	0.78378	0.74286
#06 multilearn+forgets T	0.88645	0.00000	0.55599	0.00000	0.46559	0.00001	0.94186	0.77852	0.73759
#07 multigs+multilearn T	0.61450	0.00000	0.37146	0.00003	0.24493	0.02223	0.85167	0.70968	0.64045
#08 multigs+multilearn+forgets T	0.46952	0.00000	0.29836	0.00098	0.26968	0.01154	0.81651	0.67692	0.62032
#09 multiprior T	0.88831	0.00000	0.57687	0.00000	0.47962	0.00000	0.93567	0.79730	0.74286
#10 multiprior+forgets T	0.88858	0.00000	0.55536	0.00000	0.49302	0.00000	0.90909	0.76056	0.74627
#11 multigs TE	0.59066	0.00000	0.36017	0.00006	0.21821	0.04231	0.84360	0.70213	0.63333
#12 multigs+forgets TE	0.81376	0.00000	0.47900	0.00000	0.28329	0.00784	0.00000	0.00000	0.00000
#13 multilearn TE	0.88961	0.00000	0.56588	0.00000	0.47334	0.00000	0.94118	0.78912	0.74820
#14 multilearn+forgets TE	0.88836	0.00000	0.55635	0.00000	0.46730	0.00001	0.93567	0.78378	0.74286
#15 multigs+multilearn TE	0.59291	0.00000	0.35906	0.00006	0.21142	0.04933	0.84360	0.70213	0.63333
#16 multigs+multilearn+forgets TE	0.46546	0.00000	0.30064	0.00089	0.27454	0.01007	0.81651	0.67692	0.62032
#17 multiprior TE	0.88831	0.00000	0.57687	0.00000	0.47962	0.00000	0.93567	0.79730	0.74286
#18 multiprior+forgets TE	0.88858	0.00000	0.55536	0.00000	0.49302	0.00000	0.90909	0.76056	0.74627

Table C29. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 08.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83007	0.00000	0.62161	0.00000	0.43030	0.00005	0.55072	0.73214	0.70476
#02 vanilla+forgets	0.66952	0.00000	0.39232	0.00002	0.29570	0.00699	0.00000	0.00000	0.00000
#03 multigs T	0.78634	0.00000	0.56264	0.00000	0.36931	0.00064	0.49351	0.73333	0.67257
#04 multigs+forgets T	0.60256	0.00000	0.38110	0.00003	0.27859	0.01126	0.00000	0.00000	0.00000
#05 multilearn T	0.82996	0.00000	0.62173	0.00000	0.43064	0.00005	0.55072	0.73214	0.70476
#06 multilearn+forgets T	0.65540	0.00000	0.38115	0.00003	0.28618	0.00915	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.78638	0.00000	0.56268	0.00000	0.36939	0.00064	0.49351	0.73333	0.67257
#08 multigs+multilearn+forgets T	0.70082	0.00000	0.43895	0.00000	0.32173	0.00320	0.00000	0.00000	0.00000
#09 multiprior T	0.83592	0.00000	0.51677	0.00000	0.33908	0.00183	0.79167	0.54945	0.50000
#10 multiprior+forgets T	0.60591	0.00000	0.33934	0.00022	0.25241	0.02215	0.00000	0.00000	0.00000
#11 multigs TE	0.80110	0.00000	0.56763	0.00000	0.39005	0.00029	0.51351	0.71795	0.67273
#12 multigs+forgets TE	0.74712	0.00000	0.46354	0.00000	0.33799	0.00190	0.00000	0.00000	0.00000
#13 multilearn TE	0.82995	0.00000	0.62172	0.00000	0.43063	0.00005	0.55072	0.73214	0.70476
#14 multilearn+forgets TE	0.64363	0.00000	0.36718	0.00006	0.27361	0.01287	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.80106	0.00000	0.56760	0.00000	0.38998	0.00029	0.51351	0.71795	0.67273
#16 multigs+multilearn+forgets TE	0.55568	0.00000	0.31547	0.00063	0.22607	0.04112	0.00000	0.00000	0.00000
#17 multiprior TE	0.83592	0.00000	0.51677	0.00000	0.33908	0.00183	0.79167	0.54945	0.50000
#18 multiprior+forgets TE	0.60591	0.00000	0.33934	0.00022	0.25241	0.02215	0.00000	0.00000	0.00000

Table C30. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 09.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.78479	0.00000	0.42292	0.00000	0.33096	0.00174	0.56716	0.74074	0.68831
#02 vanilla+forgets	0.72843	0.00000	0.38972	0.00001	0.27269	0.01061	0.00000	0.00000	0.00000
#03 multigs T	0.73071	0.00000	0.42156	0.00000	0.43058	0.00003	0.54286	0.73810	0.71250
#04 multigs+forgets T	0.72454	0.00000	0.38387	0.00001	0.26703	0.01241	0.00000	0.00000	0.00000
#05 multilearn T	0.78509	0.00000	0.42287	0.00000	0.33089	0.00175	0.56716	0.74074	0.68831
#06 multilearn+forgets T	0.79316	0.00000	0.44462	0.00000	0.32765	0.00195	0.59375	0.75641	0.70270
#07 multigs+multilearn T	0.73134	0.00000	0.42280	0.00000	0.43057	0.00003	0.54286	0.73810	0.71250
#08 multigs+multilearn+forgets T	0.78879	0.00000	0.41266	0.00000	0.34053	0.00125	0.60317	0.75325	0.68493
#09 multiprior T	0.83837	0.00000	0.56108	0.00000	0.47448	0.00000	0.88608	0.65421	0.64646
#10 multiprior+forgets T	0.68354	0.00000	0.36552	0.00004	0.31120	0.00335	0.00000	0.00000	0.00000
#11 multigs TE	0.74450	0.00000	0.41958	0.00000	0.38846	0.00020	0.55072	0.73494	0.70886
#12 multigs+forgets TE	0.73891	0.00000	0.39138	0.00001	0.26703	0.01241	0.00000	0.00000	0.00000
#13 multilearn TE	0.78511	0.00000	0.42286	0.00000	0.33088	0.00175	0.56716	0.74074	0.68831
#14 multilearn+forgets TE	0.72294	0.00000	0.38806	0.00001	0.27649	0.00953	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.74451	0.00000	0.41959	0.00000	0.38844	0.00020	0.55072	0.73494	0.70886
#16 multigs+multilearn+forgets TE	0.64732	0.00000	0.35358	0.00007	0.29887	0.00492	0.00000	0.00000	0.00000
#17 multiprior TE	0.83837	0.00000	0.56108	0.00000	0.47448	0.00000	0.88608	0.65421	0.64646
#18 multiprior+forgets TE	0.68354	0.00000	0.36552	0.00004	0.31120	0.00335	0.00000	0.00000	0.00000

Table C31. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 10.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.81734	0.00000	0.40813	0.00001	0.38260	0.00084	0.95135	0.68831	0.64384
#02 vanilla+forgets	0.82086	0.00000	0.41965	0.00001	0.37848	0.00096	0.95604	0.70199	0.65734
#03 multigs T	0.57816	0.00000	0.31152	0.00109	0.24489	0.03679	0.86275	0.65896	0.59394
#04 multigs+forgets T	0.57816	0.00000	0.31152	0.00109	0.24489	0.03679	0.86275	0.65896	0.59394
#05 multilearn T	0.83160	0.00000	0.41151	0.00001	0.37967	0.00092	0.95604	0.70199	0.65734
#06 multilearn+forgets T	0.83481	0.00000	0.42950	0.00000	0.38212	0.00085	0.96133	0.70667	0.66197
#07 multigs+multilearn T	0.58865	0.00000	0.32198	0.00072	0.24891	0.03371	0.86275	0.65896	0.59394
#08 multigs+multilearn+forgets T	0.57803	0.00000	0.31304	0.00103	0.24795	0.03442	0.86275	0.65896	0.59394
#09 multiprior T	0.81158	0.00000	0.40937	0.00001	0.37392	0.00112	0.95082	0.69737	0.65278
#10 multiprior+forgets T	0.83013	0.00000	0.43494	0.00000	0.42440	0.00018	0.94186	0.73759	0.70677
#11 multigs TE	0.62657	0.00000	0.33621	0.00040	0.26321	0.02446	0.87129	0.66667	0.60123
#12 multigs+forgets TE	0.61600	0.00000	0.31854	0.00083	0.25170	0.03170	0.86700	0.66279	0.59756
#13 multilearn TE	0.82853	0.00000	0.41072	0.00001	0.37811	0.00097	0.95604	0.70199	0.65734
#14 multilearn+forgets TE	0.84338	0.00000	0.42961	0.00000	0.38827	0.00069	0.96667	0.71141	0.66667
#15 multigs+multilearn TE	0.62650	0.00000	0.33624	0.00040	0.26372	0.02417	0.87129	0.66667	0.60123
#16 multigs+multilearn+forgets TE	0.61247	0.00000	0.32074	0.00076	0.25451	0.02978	0.86700	0.66279	0.59756
#17 multiprior TE	0.81158	0.00000	0.40937	0.00001	0.37392	0.00112	0.95082	0.69737	0.65278
#18 multiprior+forgets TE	0.83013	0.00000	0.43494	0.00000	0.42440	0.00018	0.94186	0.73759	0.70677

Table C32. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 11.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82945	0.00000	0.68610	0.00000	0.41324	0.00028	0.80000	0.78947	0.67925
#02 vanilla+forgets	0.82009	0.00000	0.66444	0.00000	0.54384	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.74108	0.00000	0.63154	0.00000	0.53546	0.00000	0.69725	0.76562	0.70000
#04 multigs+forgets T	0.72922	0.00000	0.59561	0.00000	0.48905	0.00001	0.67857	0.76336	0.69919
#05 multilearn T	0.82902	0.00000	0.68501	0.00000	0.41152	0.00030	0.80000	0.78947	0.67925
#06 multilearn+forgets T	0.80776	0.00000	0.69569	0.00000	0.45320	0.00006	0.78261	0.75676	0.64078
#07 multigs+multilearn T	0.73628	0.00000	0.61474	0.00000	0.53455	0.00000	0.69091	0.75969	0.69421
#08 multigs+multilearn+forgets T	0.73682	0.00000	0.61808	0.00000	0.53388	0.00000	0.68468	0.76923	0.70492
#09 multiprior T	0.86055	0.00000	0.69331	0.00000	0.46871	0.00003	0.85393	0.75926	0.66000
#10 multiprior+forgets T	0.82044	0.00000	0.66118	0.00000	0.53338	0.00000	0.00000	0.00000	0.00000
#11 multigs TE	0.73652	0.00000	0.61483	0.00000	0.53671	0.00000	0.69091	0.75969	0.69421
#12 multigs+forgets TE	0.72909	0.00000	0.59575	0.00000	0.48776	0.00001	0.67857	0.76336	0.69919
#13 multilearn TE	0.82916	0.00000	0.68533	0.00000	0.41199	0.00029	0.80000	0.78947	0.67925
#14 multilearn+forgets TE	0.82190	0.00000	0.63328	0.00000	0.43085	0.00014	0.70423	0.55556	0.48780
#15 multigs+multilearn TE	0.73597	0.00000	0.61494	0.00000	0.53707	0.00000	0.69091	0.75969	0.69421
#16 multigs+multilearn+forgets TE	0.72974	0.00000	0.59642	0.00000	0.48543	0.00001	0.67857	0.76336	0.69919
#17 multiprior TE	0.86055	0.00000	0.69331	0.00000	0.46871	0.00003	0.85393	0.75926	0.66000
#18 multiprior+forgets TE	0.82044	0.00000	0.66118	0.00000	0.53338	0.00000	0.00000	0.00000	0.00000

Table C33. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 12.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82664	0.00000	0.54551	0.00000	0.56133	0.00000	0.84783	0.68519	0.68000
#02 vanilla+forgets	0.83656	0.00000	0.58899	0.00000	0.65222	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.65191	0.00000	0.47406	0.00000	0.43431	0.00017	0.63415	0.73381	0.65649
#04 multigs+forgets T	0.61055	0.00000	0.44289	0.00000	0.47964	0.00003	0.60465	0.73103	0.65693
#05 multilearn T	0.82574	0.00000	0.54489	0.00000	0.56005	0.00000	0.84783	0.68519	0.68000
#06 multilearn+forgets T	0.83782	0.00000	0.58754	0.00000	0.64763	0.00000	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.65148	0.00000	0.47200	0.00000	0.43402	0.00017	0.63415	0.73381	0.65649
#08 multigs+multilearn+forgets T	0.62017	0.00000	0.43137	0.00001	0.47627	0.00003	0.60938	0.72222	0.64706
#09 multiprior T	0.85994	0.00000	0.60716	0.00000	0.52876	0.00000	0.88000	0.68132	0.65060
#10 multiprior+forgets T	0.81176	0.00000	0.58028	0.00000	0.63755	0.00000	0.00000	0.00000	0.00000
#11 multigs TE	0.65204	0.00000	0.46745	0.00000	0.43782	0.00015	0.62903	0.72857	0.65152
#12 multigs+forgets TE	0.59057	0.00000	0.40258	0.00004	0.37842	0.00124	0.59542	0.72109	0.64748
#13 multilearn TE	0.82620	0.00000	0.54453	0.00000	0.56034	0.00000	0.84783	0.68519	0.68000
#14 multilearn+forgets TE	0.82712	0.00000	0.58292	0.00000	0.63595	0.00000	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.64902	0.00000	0.46230	0.00000	0.43591	0.00016	0.62903	0.72857	0.65152
#16 multigs+multilearn+forgets TE	0.62597	0.00000	0.38299	0.00009	0.49488	0.00001	0.64463	0.68613	0.65116
#17 multiprior TE	0.85994	0.00000	0.60716	0.00000	0.52876	0.00000	0.88000	0.68132	0.65060
#18 multiprior+forgets TE	0.81176	0.00000	0.58028	0.00000	0.63755	0.00000	0.00000	0.00000	0.00000

Table C34. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 13.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.81802	0.00000	0.65925	0.00000	0.56700	0.00000	0.73171	0.77778	0.70000
#02 vanilla+forgets	0.78695	0.00000	0.60097	0.00000	0.55296	0.00000	0.00000	0.00000	0.00000
#03 multigs T	0.81475	0.00000	0.64389	0.00000	0.52938	0.00000	0.69767	0.78571	0.71154
#04 multigs+forgets T	0.78337	0.00000	0.59764	0.00000	0.55428	0.00000	0.00000	0.00000	0.00000
#05 multilearn T	0.81947	0.00000	0.65495	0.00000	0.56477	0.00000	0.73171	0.77778	0.70000
#06 multilearn+forgets T	0.78933	0.00000	0.60233	0.00000	0.55309	0.00000	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.81609	0.00000	0.66427	0.00000	0.56661	0.00000	0.71429	0.80000	0.72549
#08 multigs+multilearn+forgets T	0.78716	0.00000	0.59975	0.00000	0.55495	0.00000	0.00000	0.00000	0.00000
#09 multiprior T	0.83542	0.00000	0.66985	0.00000	0.54344	0.00000	0.74074	0.76636	0.70707
#10 multiprior+forgets T	0.74465	0.00000	0.57353	0.00000	0.53698	0.00000	0.00000	0.00000	0.00000
#11 multigs TE	0.75323	0.00000	0.58710	0.00000	0.48562	0.00002	0.60606	0.78400	0.70085
#12 multigs+forgets TE	0.69653	0.00000	0.46641	0.00000	0.51682	0.00000	0.00000	0.00000	0.00000
#13 multilearn TE	0.81985	0.00000	0.65305	0.00000	0.56208	0.00000	0.73171	0.77778	0.70000
#14 multilearn+forgets TE	0.79417	0.00000	0.60649	0.00000	0.55265	0.00000	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.75458	0.00000	0.58837	0.00000	0.48446	0.00002	0.60606	0.78400	0.70085
#16 multigs+multilearn+forgets TE	0.76982	0.00000	0.56892	0.00000	0.46753	0.00004	0.68182	0.75439	0.66038
#17 multiprior TE	0.83542	0.00000	0.66985	0.00000	0.54344	0.00000	0.74074	0.76636	0.70707
#18 multiprior+forgets TE	0.74465	0.00000	0.57353	0.00000	0.53698	0.00000	0.00000	0.00000	0.00000

Table C35. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 14.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82146	0.00000	0.58523	0.00000	0.46517	0.00004	0.84800	0.74419	0.67227
#02 vanilla+forgets	0.75177	0.00000	0.61163	0.00000	0.50664	0.00001	0.00000	0.00000	0.00000
#03 multigs T	0.68558	0.00000	0.44403	0.00000	0.33967	0.00351	0.74648	0.72603	0.66176
#04 multigs+forgets T	0.61925	0.00000	0.39980	0.00005	0.34644	0.00287	0.72109	0.71523	0.65248
#05 multilearn T	0.82279	0.00000	0.58135	0.00000	0.45740	0.00005	0.84800	0.74419	0.67227
#06 multilearn+forgets T	0.85183	0.00000	0.68563	0.00000	0.53438	0.00000	0.75862	0.65934	0.59259
#07 multigs+multilearn T	0.68563	0.00000	0.44407	0.00000	0.33994	0.00348	0.74648	0.72603	0.66176
#08 multigs+multilearn+forgets T	0.65037	0.00000	0.42178	0.00002	0.35468	0.00224	0.73611	0.71622	0.65217
#09 multiprior T	0.83688	0.00000	0.62472	0.00000	0.50490	0.00001	0.86885	0.76190	0.70690
#10 multiprior+forgets T	0.72048	0.00000	0.59130	0.00000	0.49568	0.00001	0.00000	0.00000	0.00000
#11 multigs TE	0.68557	0.00000	0.44404	0.00000	0.33968	0.00351	0.74648	0.72603	0.66176
#12 multigs+forgets TE	0.62119	0.00000	0.40146	0.00004	0.34877	0.00268	0.72109	0.71523	0.65248
#13 multilearn TE	0.82340	0.00000	0.57858	0.00000	0.45544	0.00006	0.84800	0.74419	0.67227
#14 multilearn+forgets TE	0.85231	0.00000	0.66771	0.00000	0.59848	0.00000	0.85417	0.70000	0.71111
#15 multigs+multilearn TE	0.68540	0.00000	0.44453	0.00000	0.34027	0.00345	0.74648	0.72603	0.66176
#16 multigs+multilearn+forgets TE	0.63694	0.00000	0.42418	0.00001	0.35968	0.00191	0.72603	0.72000	0.65714
#17 multiprior TE	0.83688	0.00000	0.62472	0.00000	0.50490	0.00001	0.86885	0.76190	0.70690
#18 multiprior+forgets TE	0.72048	0.00000	0.59130	0.00000	0.49568	0.00001	0.00000	0.00000	0.00000

Table C36. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 15.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.85194	0.00000	0.45365	0.00000	0.32196	0.00457	0.89933	0.73381	0.66165
#02 vanilla+forgets	0.84337	0.00000	0.43058	0.00000	0.29150	0.01062	0.89933	0.71942	0.66165
#03 multigs T	0.75403	0.00000	0.43404	0.00000	0.17257	0.13605	0.83750	0.69333	0.65278
#04 multigs+forgets T	0.72103	0.00000	0.41494	0.00001	0.16479	0.15487	0.82716	0.68421	0.64384
#05 multilearn T	0.86944	0.00000	0.45003	0.00000	0.31065	0.00631	0.90541	0.73913	0.66667
#06 multilearn+forgets T	0.85550	0.00000	0.43170	0.00000	0.33325	0.00326	0.90541	0.72464	0.66667
#07 multigs+multilearn T	0.75417	0.00000	0.43391	0.00000	0.17256	0.13605	0.83750	0.69333	0.65278
#08 multigs+multilearn+forgets T	0.73730	0.00000	0.43706	0.00000	0.16991	0.14227	0.83230	0.68874	0.64828
#09 multiprior T	0.84933	0.00000	0.42519	0.00001	0.33192	0.00340	0.90541	0.72464	0.66667
#10 multiprior+forgets T	0.82674	0.00000	0.40597	0.00002	0.29654	0.00929	0.88742	0.70922	0.66667
#11 multigs TE	0.75342	0.00000	0.43197	0.00000	0.17916	0.12149	0.84277	0.69799	0.64336
#12 multigs+forgets TE	0.75280	0.00000	0.43385	0.00000	0.17162	0.13824	0.83750	0.69333	0.65278
#13 multilearn TE	0.86803	0.00000	0.44950	0.00000	0.31207	0.00606	0.90541	0.73913	0.66667
#14 multilearn+forgets TE	0.87358	0.00000	0.43666	0.00000	0.31987	0.00485	0.92414	0.72593	0.65116
#15 multigs+multilearn TE	0.75285	0.00000	0.43214	0.00000	0.17724	0.12559	0.84277	0.69799	0.64336
#16 multigs+multilearn+forgets TE	0.74712	0.00000	0.43693	0.00000	0.17181	0.13781	0.83230	0.68874	0.64828
#17 multiprior TE	0.84933	0.00000	0.42519	0.00001	0.33192	0.00340	0.90541	0.72464	0.66667
#18 multiprior+forgets TE	0.82674	0.00000	0.40597	0.00002	0.29654	0.00929	0.88742	0.70922	0.66667

Table C37. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 16.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83085	0.00000	0.51222	0.00000	0.32629	0.00402	0.69565	0.57471	0.46914
#02 vanilla+forgets	0.86014	0.00000	0.46146	0.00000	0.45253	0.00004	0.00000	0.00000	0.00000
#03 multigs T	0.70127	0.00000	0.52881	0.00000	0.29969	0.00854	0.48485	0.72897	0.63366
#04 multigs+forgets T	0.81084	0.00000	0.44519	0.00000	0.41942	0.00016	0.00000	0.00000	0.00000
#05 multilearn T	0.83108	0.00000	0.51230	0.00000	0.32659	0.00398	0.69565	0.57471	0.46914
#06 multilearn+forgets T	0.85214	0.00000	0.45093	0.00000	0.44570	0.00005	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.70147	0.00000	0.52871	0.00000	0.29907	0.00868	0.48485	0.72897	0.63366
#08 multigs+multilearn+forgets T	0.81449	0.00000	0.44995	0.00000	0.42399	0.00014	0.00000	0.00000	0.00000
#09 multiprior T	0.85251	0.00000	0.47221	0.00000	0.27794	0.01506	0.75000	0.49383	0.40000
#10 multiprior+forgets T	0.85029	0.00000	0.46298	0.00000	0.44618	0.00005	0.00000	0.00000	0.00000
#11 multigs TE	0.69395	0.00000	0.53190	0.00000	0.31814	0.00510	0.47059	0.73394	0.66019
#12 multigs+forgets TE	0.77203	0.00000	0.44456	0.00000	0.41635	0.00018	0.00000	0.00000	0.00000
#13 multilearn TE	0.83105	0.00000	0.51229	0.00000	0.32656	0.00399	0.69565	0.57471	0.46914
#14 multilearn+forgets TE	0.84968	0.00000	0.45185	0.00000	0.44286	0.00006	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.69386	0.00000	0.53205	0.00000	0.31821	0.00509	0.47059	0.73394	0.66019
#16 multigs+multilearn+forgets TE	0.77955	0.00000	0.44171	0.00000	0.40761	0.00026	0.00000	0.00000	0.00000
#17 multiprior TE	0.85251	0.00000	0.47221	0.00000	0.27794	0.01506	0.75000	0.49383	0.40000
#18 multiprior+forgets TE	0.85029	0.00000	0.46298	0.00000	0.44618	0.00005	0.00000	0.00000	0.00000

Table C38. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 17.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.83753	0.00000	0.41707	0.00002	0.24575	0.03611	0.93243	0.76119	0.70769
#02 vanilla+forgets	0.80580	0.00000	0.40168	0.00005	0.17942	0.12880	0.93333	0.75000	0.69697
#03 multigs T	0.58685	0.00000	0.32945	0.00105	0.26490	0.02352	0.84337	0.71053	0.68919
#04 multigs+forgets T	0.50943	0.00000	0.28422	0.00501	0.22691	0.05355	0.83333	0.70130	0.68000
#05 multilearn T	0.83772	0.00000	0.41754	0.00002	0.24718	0.03501	0.93243	0.76119	0.70769
#06 multilearn+forgets T	0.83911	0.00000	0.39713	0.00006	0.24122	0.03980	0.92414	0.74809	0.70866
#07 multigs+multilearn T	0.58788	0.00000	0.32588	0.00119	0.27297	0.01946	0.84337	0.71053	0.68919
#08 multigs+multilearn+forgets T	0.24305	0.01165	0.14679	0.15354	0.18664	0.11386	0.80000	0.69565	0.66242
#09 multiprior T	0.88136	0.00000	0.41984	0.00002	0.22333	0.05753	0.94203	0.74194	0.70000
#10 multiprior+forgets T	0.87796	0.00000	0.42629	0.00001	0.22472	0.05595	0.94964	0.73600	0.69421
#11 multigs TE	0.65772	0.00000	0.32437	0.00126	0.15673	0.18544	0.86420	0.71622	0.66667
#12 multigs+forgets TE	0.62216	0.00000	0.29139	0.00397	0.20548	0.08116	0.85366	0.70667	0.67123
#13 multilearn TE	0.83769	0.00000	0.41739	0.00002	0.24650	0.03553	0.93243	0.76119	0.70769
#14 multilearn+forgets TE	0.82476	0.00000	0.39813	0.00006	0.21916	0.06248	0.91781	0.74242	0.70312
#15 multigs+multilearn TE	0.65827	0.00000	0.32398	0.00128	0.15714	0.18427	0.86420	0.71622	0.66667
#16 multigs+multilearn+forgets TE	0.69863	0.00000	0.37952	0.00014	0.26559	0.02315	0.87500	0.72603	0.67606
#17 multiprior TE	0.88136	0.00000	0.41984	0.00002	0.22333	0.05753	0.94203	0.74194	0.70000
#18 multiprior+forgets TE	0.87796	0.00000	0.42629	0.00001	0.22472	0.05595	0.94964	0.73600	0.69421

Table C39. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 19.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.82474	0.00000	0.39850	0.00005	0.33785	0.00303	0.70270	0.66667	0.65347
#02 vanilla+forgets	0.74324	0.00000	0.37620	0.00013	0.42591	0.00014	0.00000	0.00000	0.00000
#03 multigs T	0.72458	0.00000	0.44061	0.00001	0.35027	0.00206	0.59770	0.74380	0.70175
#04 multigs+forgets T	0.76029	0.00000	0.38351	0.00010	0.44279	0.00007	0.00000	0.00000	0.00000
#05 multilearn T	0.82438	0.00000	0.39895	0.00005	0.33757	0.00306	0.70270	0.66667	0.65347
#06 multilearn+forgets T	0.76506	0.00000	0.38868	0.00008	0.43811	0.00008	0.00000	0.00000	0.00000
#07 multigs+multilearn T	0.72534	0.00000	0.44090	0.00001	0.35011	0.00208	0.59770	0.74380	0.70175
#08 multigs+multilearn+forgets T	0.81002	0.00000	0.41876	0.00002	0.47358	0.00002	0.00000	0.00000	0.00000
#09 multiprior T	0.84748	0.00000	0.38240	0.00010	0.29782	0.00946	0.77612	0.63366	0.59574
#10 multiprior+forgets T	0.71556	0.00000	0.36240	0.00025	0.41017	0.00026	0.00000	0.00000	0.00000
#11 multigs TE	0.72476	0.00000	0.44086	0.00001	0.35037	0.00206	0.59770	0.74380	0.70175
#12 multigs+forgets TE	0.81740	0.00000	0.42439	0.00001	0.47600	0.00002	0.00000	0.00000	0.00000
#13 multilearn TE	0.82440	0.00000	0.39892	0.00005	0.33758	0.00306	0.70270	0.66667	0.65347
#14 multilearn+forgets TE	0.74128	0.00000	0.38012	0.00011	0.42569	0.00014	0.00000	0.00000	0.00000
#15 multigs+multilearn TE	0.72448	0.00000	0.44054	0.00001	0.35028	0.00206	0.59770	0.74380	0.70175
#16 multigs+multilearn+forgets TE	0.78478	0.00000	0.47831	0.00000	0.33880	0.00295	0.70270	0.70370	0.67327
#17 multiprior TE	0.84748	0.00000	0.38240	0.00010	0.29782	0.00946	0.77612	0.63366	0.59574
#18 multiprior+forgets TE	0.71556	0.00000	0.36240	0.00025	0.41017	0.00026	0.00000	0.00000	0.00000

Table C40. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 21.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.87205	0.00000	0.35631	0.00026	0.38317	0.00045	0.92308	0.75556	0.74419
#02 vanilla+forgets	0.89615	0.00000	0.34930	0.00034	0.35505	0.00123	0.92958	0.74627	0.73438
#03 multigs T	0.83189	0.00000	0.32833	0.00080	0.36899	0.00076	0.89189	0.74286	0.71642
#04 multigs+forgets T	0.81532	0.00000	0.32500	0.00091	0.39060	0.00034	0.88591	0.73759	0.72593
#05 multilearn T	0.87046	0.00000	0.35590	0.00026	0.38334	0.00045	0.92308	0.75556	0.74419
#06 multilearn+forgets T	0.88969	0.00000	0.34183	0.00047	0.35189	0.00137	0.92958	0.74627	0.73438
#07 multigs+multilearn T	0.83189	0.00000	0.32833	0.00080	0.36898	0.00076	0.89189	0.74286	0.71642
#08 multigs+multilearn+forgets T	0.81819	0.00000	0.32971	0.00076	0.39805	0.00026	0.88591	0.73759	0.72593
#09 multiprior T	0.91342	0.00000	0.33451	0.00063	0.31541	0.00437	0.96350	0.72868	0.69919
#10 multiprior+forgets T	0.92570	0.00000	0.36402	0.00018	0.26870	0.01595	0.97744	0.72000	0.67227
#11 multigs TE	0.83193	0.00000	0.32835	0.00080	0.36897	0.00076	0.89189	0.74286	0.71642
#12 multigs+forgets TE	0.81536	0.00000	0.32505	0.00091	0.39060	0.00034	0.88591	0.73759	0.72593
#13 multilearn TE	0.87428	0.00000	0.38035	0.00009	0.38340	0.00045	0.92958	0.76119	0.75000
#14 multilearn+forgets TE	0.87670	0.00000	0.35906	0.00023	0.35793	0.00112	0.94203	0.73846	0.72581
#15 multigs+multilearn TE	0.83193	0.00000	0.32835	0.00080	0.36897	0.00076	0.89189	0.74286	0.71642
#16 multigs+multilearn+forgets TE	0.85635	0.00000	0.31693	0.00124	0.36555	0.00086	0.90278	0.73529	0.72308
#17 multiprior TE	0.91342	0.00000	0.33451	0.00063	0.31541	0.00437	0.96350	0.72868	0.69919
#18 multiprior+forgets TE	0.92570	0.00000	0.36402	0.00018	0.26870	0.01595	0.97744	0.72000	0.67227

Table C41. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 23.

BKT model	Correlation						Classification		
	Formative		Summative - Midterm		Summative - Final		Formative	Summative - Midterm	Summative - Final
	Pearson r	p	Pearson r	p	Pearson r	p	F_1	F_1	F_1
#01 vanilla	0.84959	0.00000	0.35050	0.00101	0.39715	0.00061	0.91228	0.73504	0.74545
#02 vanilla+forgets	0.85260	0.00000	0.35789	0.00077	0.39747	0.00060	0.92035	0.74138	0.75229
#03 multigs T	0.78646	0.00000	0.30360	0.00473	0.38623	0.00088	0.88136	0.74380	0.73684
#04 multigs+forgets T	0.78722	0.00000	0.30294	0.00483	0.38750	0.00084	0.88136	0.74380	0.73684
#05 multilearn T	0.85043	0.00000	0.35238	0.00094	0.39561	0.00064	0.91228	0.73504	0.74545
#06 multilearn+forgets T	0.85343	0.00000	0.36214	0.00066	0.39300	0.00070	0.92035	0.74138	0.75229
#07 multigs+multilearn T	0.78816	0.00000	0.30754	0.00420	0.38345	0.00096	0.88136	0.74380	0.73684
#08 multigs+multilearn+forgets T	0.80001	0.00000	0.32946	0.00208	0.38413	0.00094	0.88889	0.75000	0.74336
#09 multiprior T	0.87346	0.00000	0.31623	0.00319	0.39561	0.00064	0.93694	0.73684	0.76636
#10 multiprior+forgets T	0.88333	0.00000	0.34082	0.00141	0.40136	0.00052	0.93694	0.73684	0.76636
#11 multigs TE	0.78477	0.00000	0.30050	0.00520	0.38804	0.00083	0.88136	0.74380	0.73684
#12 multigs+forgets TE	0.78718	0.00000	0.30283	0.00485	0.38752	0.00084	0.88136	0.74380	0.73684
#13 multilearn TE	0.84961	0.00000	0.35055	0.00101	0.39711	0.00061	0.91228	0.73504	0.74545
#14 multilearn+forgets TE	0.85336	0.00000	0.36126	0.00068	0.39664	0.00062	0.92035	0.74138	0.75229
#15 multigs+multilearn TE	0.78568	0.00000	0.30186	0.00499	0.38721	0.00085	0.88136	0.74380	0.73684
#16 multigs+multilearn+forgets TE	0.78968	0.00000	0.30534	0.00449	0.39019	0.00077	0.88136	0.74380	0.73684
#17 multiprior TE	0.87346	0.00000	0.31623	0.00319	0.39561	0.00064	0.93694	0.73684	0.76636
#18 multiprior+forgets TE	0.88333	0.00000	0.34082	0.00141	0.40136	0.00052	0.93694	0.73684	0.76636

Table C42. The adaptive BKT mastery probability and the overall student performance in formative and summative testing (complete dataset) – DKC 25.

APPENDIX D

Answer opportunities analysis – DKC 01-DKC 25

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	13.32000	7.59702	13.75714	7.46088	3
#02 vanilla+forgets	13.68000	7.53963	14.21429	7.36402	3
#03 multigs T	8.95000	8.41040	9.50000	8.61537	1
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	13.01000	8.07352	13.47143	7.93948	2
#06 multilearn+forgets T	13.97000	7.44604	14.52857	7.28453	3
#07 multigs+multilearn T	8.95000	8.41040	9.50000	8.61537	1
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	14.28000	6.88444	14.82857	6.72864	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	8.95000	8.41040	9.50000	8.61537	1
#12 multigs+forgets TE	8.93000	8.41266	9.50000	8.61537	1
#13 multilearn TE	13.01000	8.07352	13.47143	7.93948	2
#14 multilearn+forgets TE	14.13000	7.39732	14.71429	7.24147	3
#15 multigs+multilearn TE	8.95000	8.41040	9.50000	8.61537	1
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	14.28000	6.88444	14.82857	6.72864	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D1. Answer opportunities analysis – DKC 01.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	11.07619	7.78274	11.60274	7.73473	3
#02 vanilla+forgets	11.37143	7.86821	11.95890	7.80569	3
#03 multigs T	5.20952	6.69773	5.39726	6.76950	1
#04 multigs+forgets T	4.40952	6.02350	4.34247	5.83147	1
#05 multilearn T	11.42857	7.91403	12.04110	7.86596	3
#06 multilearn+forgets T	13.07619	7.26252	13.63014	7.20244	4
#07 multigs+multilearn T	5.20952	6.69773	5.39726	6.76950	1
#08 multigs+multilearn+forgets T	5.09524	6.63256	5.19178	6.54611	1
#09 multiprior T	14.16190	6.11763	14.47945	5.93041	6
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	5.26667	6.76852	5.47945	6.86802	1
#12 multigs+forgets TE	4.40952	6.02350	4.34247	5.83147	1
#13 multilearn TE	11.29524	7.82614	11.89041	7.78275	3
#14 multilearn+forgets TE	11.89524	7.77190	12.50685	7.74081	3
#15 multigs+multilearn TE	5.20952	6.69773	5.39726	6.76950	1
#16 multigs+multilearn+forgets TE	5.02857	6.54196	5.16438	6.54687	1
#17 multiprior TE	14.16190	6.11763	14.47945	5.93041	6
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D2. Answer opportunities analysis – DKC 02.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	9.80682	7.71751	10.27869	7.74840	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	6.68182	7.36963	7.32787	7.57787	1
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	9.87500	7.78270	10.37705	7.83616	2
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	6.68182	7.36963	7.32787	7.57787	1
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	16.76136	5.15059	16.59016	5.20377	6
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	6.76136	7.45561	7.44262	7.68879	1
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	9.80682	7.71751	10.27869	7.74840	2
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	6.72727	7.43976	7.39344	7.67089	1
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	16.76136	5.15059	16.59016	5.20377	6
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D3. Answer opportunities analysis – DKC 03.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	10.21795	8.18241	11.50000	8.21354	2
#02 vanilla+forgets	10.20513	8.17878	11.50000	8.21354	2
#03 multigs T	8.10256	8.32550	9.42593	8.50882	1
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	10.29487	8.21166	11.61111	8.23801	2
#06 multilearn+forgets T	10.42308	8.14562	11.77778	8.12326	2
#07 multigs+multilearn T	8.03846	8.28305	9.33333	8.46302	1
#08 multigs+multilearn+forgets T	8.03846	8.28305	9.33333	8.46302	1
#09 multiprior T	11.20513	7.66906	12.40741	7.66398	3
#10 multiprior+forgets T	11.53846	7.67809	12.66667	7.61825	3
#11 multigs TE	8.03846	8.28305	9.33333	8.46302	1
#12 multigs+forgets TE	7.98718	8.30974	9.27778	8.49954	1
#13 multilearn TE	10.24359	8.19437	11.53704	8.22493	2
#14 multilearn+forgets TE	10.84615	8.17833	12.07407	8.11672	2
#15 multigs+multilearn TE	8.03846	8.28305	9.33333	8.46302	1
#16 multigs+multilearn+forgets TE	7.98718	8.30974	9.27778	8.49954	1
#17 multiprior TE	11.20513	7.66906	12.40741	7.66398	3
#18 multiprior+forgets TE	11.53846	7.67809	12.66667	7.61825	3

Table D4. Answer opportunities analysis – DKC 04.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	15.24324	6.45492	15.19231	6.33397	3
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	13.52703	7.38607	13.38462	7.39145	2
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	15.24324	6.45492	15.19231	6.33397	3
#06 multilearn+forgets T	18.87838	4.06439	18.73077	4.26162	2
#07 multigs+multilearn T	13.50000	7.38427	13.34615	7.38808	2
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	16.22973	5.68957	16.25000	5.54438	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	14.10811	6.98642	14.07692	6.81649	2
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	15.24324	6.45492	15.19231	6.33397	3
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	14.10811	6.98642	14.07692	6.81649	2
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	16.22973	5.68957	16.25000	5.54438	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D5. Answer opportunities analysis – DKC 05.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	10.33766	8.94076	10.48148	8.89295	1
#02 vanilla+forgets	10.33766	8.94076	10.48148	8.89295	1
#03 multigs T	10.31169	8.94907	10.46296	8.89077	1
#04 multigs+forgets T	9.83117	8.88731	10.07407	8.99654	1
#05 multilearn T	10.75325	8.95479	10.55556	8.83746	1
#06 multilearn+forgets T	10.35065	8.93804	10.50000	8.88873	1
#07 multigs+multilearn T	10.33766	8.94076	10.48148	8.89295	1
#08 multigs+multilearn+forgets T	9.83117	8.88731	10.07407	8.99654	1
#09 multiprior T	11.97403	8.01310	11.79630	7.92745	2
#10 multiprior+forgets T	11.80519	8.18059	11.59259	8.12524	2
#11 multigs TE	10.63636	8.89546	10.53704	8.84182	1
#12 multigs+forgets TE	10.31169	8.94907	10.46296	8.89077	1
#13 multilearn TE	10.63636	8.89546	10.53704	8.84182	1
#14 multilearn+forgets TE	10.33766	8.94076	10.48148	8.89295	1
#15 multigs+multilearn TE	10.33766	8.94076	10.48148	8.89295	1
#16 multigs+multilearn+forgets TE	10.25974	8.97863	10.38889	8.93460	1
#17 multiprior TE	11.97403	8.01310	11.79630	7.92745	2
#18 multiprior+forgets TE	11.80519	8.18059	11.59259	8.12524	2

Table D6. Answer opportunities analysis – DKC 06.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	17.13889	6.09850	17.42000	6.08139	2
#02 vanilla+forgets	17.19444	6.05989	17.50000	6.02122	2
#03 multigs T	17.06944	6.15830	17.40000	6.07437	2
#04 multigs+forgets T	17.08333	6.16384	17.42000	6.08139	2
#05 multilearn T	17.13889	6.09850	17.42000	6.08139	2
#06 multilearn+forgets T	17.44444	5.78947	17.86000	5.59887	20
#07 multigs+multilearn T	17.06944	6.15830	17.40000	6.07437	2
#08 multigs+multilearn+forgets T	17.27778	5.96049	17.76000	5.66248	20
#09 multiprior T	17.91667	5.00352	18.06000	5.16033	3
#10 multiprior+forgets T	17.98611	4.92334	18.06000	5.16033	3
#11 multigs TE	17.06944	6.15830	17.40000	6.07437	2
#12 multigs+forgets TE	17.02778	6.21668	17.40000	6.07437	2
#13 multilearn TE	17.13889	6.09850	17.42000	6.08139	2
#14 multilearn+forgets TE	17.44444	5.78947	17.86000	5.59887	20
#15 multigs+multilearn TE	17.06944	6.15830	17.40000	6.07437	2
#16 multigs+multilearn+forgets TE	17.27778	5.96049	17.76000	5.66248	20
#17 multiprior TE	17.91667	5.00352	18.06000	5.16033	3
#18 multiprior+forgets TE	17.98611	4.92334	18.06000	5.16033	3

Table D7. Answer opportunities analysis – DKC 07.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	11.72464	7.43023	11.85417	7.58528	3
#02 vanilla+forgets	12.29710	7.07871	12.26042	7.17928	4
#03 multigs T	4.57246	6.22555	4.07292	5.68515	1
#04 multigs+forgets T	3.07971	4.63573	2.54167	3.64740	1
#05 multilearn T	12.02174	7.47108	11.94792	7.57888	3
#06 multilearn+forgets T	12.49275	7.05375	12.40625	7.16362	4
#07 multigs+multilearn T	4.71014	6.38338	4.23958	5.93428	1
#08 multigs+multilearn+forgets T	3.20290	4.78007	2.66667	3.82421	1
#09 multiprior T	13.92754	5.85922	13.70833	6.02961	6
#10 multiprior+forgets T	14.44203	5.91549	14.08333	6.07266	6
#11 multigs TE	4.15217	6.11436	3.80208	5.68585	1
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	12.10145	7.49140	12.01042	7.60816	3
#14 multilearn+forgets TE	12.17391	7.43135	12.07292	7.55227	3
#15 multigs+multilearn TE	4.15217	6.11436	3.80208	5.68585	1
#16 multigs+multilearn+forgets TE	3.20290	4.78007	2.66667	3.82421	1
#17 multiprior TE	13.92754	5.85922	13.70833	6.02961	6
#18 multiprior+forgets TE	14.44203	5.91549	14.08333	6.07266	6

Table D8. Answer opportunities analysis – DKC 08.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	14.31538	7.45013	14.70330	7.34922	3
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	12.70769	8.19098	13.20879	8.08224	2
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	14.31538	7.45013	14.70330	7.34922	3
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	12.70769	8.19098	13.20879	8.08224	2
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	17.35385	5.05913	17.67033	4.81446	6
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	13.31538	8.00197	13.76923	7.96810	2
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	14.31538	7.45013	14.70330	7.34922	3
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	13.31538	8.00197	13.76923	7.96810	2
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	17.35385	5.05913	17.67033	4.81446	6
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D9. Answer opportunities analysis – DKC 09.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	8.55396	7.78253	8.38144	7.69530	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	6.89928	8.03281	6.78351	7.92047	1
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	8.55396	7.78253	8.38144	7.69530	2
#06 multilearn+forgets T	9.46043	8.02195	9.45361	8.02862	20
#07 multigs+multilearn T	6.89209	8.02865	6.77320	7.91426	1
#08 multigs+multilearn+forgets T	9.83453	7.97332	9.78351	7.93886	20
#09 multiprior T	17.35971	4.49478	17.38144	4.60173	7
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	7.12230	8.18222	7.10309	8.14003	1
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	8.55396	7.78253	8.38144	7.69530	2
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	7.12230	8.18222	7.10309	8.14003	1
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	17.35971	4.49478	17.38144	4.60173	7
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D10. Answer opportunities analysis – DKC 10.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	8.33594	7.60571	8.29213	7.46538	2
#02 vanilla+forgets	9.60938	7.21791	9.51685	7.01640	3
#03 multigs T	4.03906	5.75812	3.97753	5.69086	1
#04 multigs+forgets T	4.03906	5.75812	3.97753	5.69086	1
#05 multilearn T	8.65625	7.75657	8.55056	7.56789	2
#06 multilearn+forgets T	10.03125	7.29679	10.07865	7.11948	3
#07 multigs+multilearn T	4.16406	5.98131	4.07865	5.86830	1
#08 multigs+multilearn+forgets T	4.03906	5.75812	3.97753	5.69086	1
#09 multiprior T	10.42969	6.60909	10.30337	6.37539	4
#10 multiprior+forgets T	12.12500	6.53398	12.08989	6.39183	5
#11 multigs TE	4.48438	6.38031	4.35955	6.19979	1
#12 multigs+forgets TE	4.33594	6.18135	4.15730	5.92548	1
#13 multilearn TE	8.63281	7.75092	8.51685	7.55903	2
#14 multilearn+forgets TE	10.01563	7.31301	10.06742	7.13633	3
#15 multigs+multilearn TE	4.48438	6.38031	4.35955	6.19979	1
#16 multigs+multilearn+forgets TE	4.31250	6.17175	4.14607	5.92098	1
#17 multiprior TE	10.42969	6.60909	10.30337	6.37539	4
#18 multiprior+forgets TE	12.12500	6.53398	12.08989	6.39183	5

Table D11. Answer opportunities analysis – DKC 11.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	12.39844	8.54938	12.33708	8.54659	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	10.08594	9.00658	10.17978	9.07990	1
#04 multigs+forgets T	9.71094	8.97780	9.77528	9.03372	1
#05 multilearn T	12.39844	8.54938	12.33708	8.54659	2
#06 multilearn+forgets T	13.64844	7.78004	13.68539	7.77904	20
#07 multigs+multilearn T	9.88281	8.96065	9.98876	9.03339	1
#08 multigs+multilearn+forgets T	9.82813	8.99878	9.91011	9.08875	1
#09 multiprior T	14.27344	7.13836	14.19101	7.12371	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	9.87500	8.96186	9.97753	9.03526	1
#12 multigs+forgets TE	9.71094	8.97780	9.77528	9.03372	1
#13 multilearn TE	12.39844	8.54938	12.33708	8.54659	2
#14 multilearn+forgets TE	16.13281	6.73032	16.31461	6.64625	20
#15 multigs+multilearn TE	9.87500	8.96186	9.97753	9.03526	1
#16 multigs+multilearn+forgets TE	9.71094	8.97780	9.77528	9.03372	1
#17 multiprior TE	14.27344	7.13836	14.19101	7.12371	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D12. Answer opportunities analysis – DKC 12.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	13.73554	7.46075	13.74118	7.39233	3
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	8.18182	8.31064	7.98824	8.45928	1
#04 multigs+forgets T	7.30579	8.02895	7.17647	8.19578	1
#05 multilearn T	13.79339	7.47542	13.75294	7.39193	3
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	8.18182	8.31064	7.98824	8.45928	1
#08 multigs+multilearn+forgets T	7.49587	8.07684	7.35294	8.29186	1
#09 multiprior T	16.55372	5.61687	16.51765	5.64128	6
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	8.04132	8.22638	7.94118	8.45345	1
#12 multigs+forgets TE	6.70248	7.80240	6.65882	7.88240	1
#13 multilearn TE	13.73554	7.46075	13.74118	7.39233	3
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	8.09091	8.26942	7.94118	8.45345	1
#16 multigs+multilearn+forgets TE	8.52066	8.22304	8.56471	8.25551	1
#17 multiprior TE	16.55372	5.61687	16.51765	5.64128	6
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D13. Answer opportunities analysis – DKC 13.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	13.04065	8.21973	13.45349	8.15678	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	12.94309	8.18866	13.38372	8.14309	2
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	13.04065	8.21973	13.45349	8.15678	2
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	13.02439	8.20782	13.43023	8.14073	2
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	14.56911	6.97429	14.77907	6.96952	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	11.00813	8.22481	11.55814	8.29827	2
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	13.04065	8.21973	13.45349	8.15678	2
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	11.00813	8.22481	11.55814	8.29827	2
#16 multigs+multilearn+forgets TE	12.62602	8.07384	12.97674	8.08481	2
#17 multiprior TE	14.56911	6.97429	14.77907	6.96952	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D14. Answer opportunities analysis – DKC 14.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	10.12069	8.18817	10.62963	8.24840	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	6.46552	7.80795	6.64198	7.91250	1
#04 multigs+forgets T	5.73276	7.37016	5.95062	7.50317	1
#05 multilearn T	10.17241	8.21587	10.70370	8.28318	2
#06 multilearn+forgets T	16.18966	6.34914	16.55556	6.16036	20
#07 multigs+multilearn T	6.46552	7.80795	6.64198	7.91250	1
#08 multigs+multilearn+forgets T	6.03448	7.63736	6.20988	7.69532	1
#09 multiprior T	11.87931	7.10561	12.38272	7.19473	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	6.46552	7.80795	6.64198	7.91250	1
#12 multigs+forgets TE	5.73276	7.37016	5.95062	7.50317	1
#13 multilearn TE	10.18103	8.21938	10.70370	8.28318	2
#14 multilearn+forgets TE	15.50862	6.30320	15.82716	6.28250	5
#15 multigs+multilearn TE	6.46552	7.80795	6.64198	7.91250	1
#16 multigs+multilearn+forgets TE	5.89655	7.50116	6.12346	7.64425	1
#17 multiprior TE	11.87931	7.10561	12.38272	7.19473	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D15. Answer opportunities analysis – DKC 15.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	9.00000	8.20804	9.70588	8.46256	2
#02 vanilla+forgets	9.31148	8.07514	9.78824	8.28811	2
#03 multigs T	6.29508	7.99244	6.85882	8.18558	1
#04 multigs+forgets T	6.01639	7.78533	6.64706	8.05739	1
#05 multilearn T	9.37705	8.14063	10.09412	8.41289	2
#06 multilearn+forgets T	9.61475	8.05708	10.07059	8.30345	2
#07 multigs+multilearn T	6.29508	7.99244	6.85882	8.18558	1
#08 multigs+multilearn+forgets T	6.15574	7.89499	6.82353	8.18997	1
#09 multiprior T	10.03279	7.62655	10.48235	7.85677	3
#10 multiprior+forgets T	9.70492	7.54676	10.17647	7.74163	3
#11 multigs TE	6.37705	8.00860	6.95294	8.20555	1
#12 multigs+forgets TE	6.29508	7.99244	6.85882	8.18558	1
#13 multilearn TE	9.37705	8.14063	10.09412	8.41289	2
#14 multilearn+forgets TE	9.60656	8.17531	10.11765	8.43379	2
#15 multigs+multilearn TE	6.36885	8.00331	6.94118	8.19894	1
#16 multigs+multilearn+forgets TE	6.17213	7.89151	6.84706	8.18318	1
#17 multiprior TE	10.03279	7.62655	10.48235	7.85677	3
#18 multiprior+forgets TE	9.70492	7.54676	10.17647	7.74163	3

Table D16. Answer opportunities analysis – DKC 16.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	15.86885	7.13269	16.15294	6.99065	3
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	13.36066	8.15862	13.60000	8.06698	2
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	15.86885	7.13269	16.15294	6.99065	3
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	13.36066	8.15862	13.60000	8.06698	2
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	17.16393	5.65007	17.52941	5.40632	5
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	13.09836	8.20140	13.38824	8.15781	2
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	15.86885	7.13269	16.15294	6.99065	3
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	13.09836	8.20140	13.38824	8.15781	2
#16 multigs+multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#17 multiprior TE	17.16393	5.65007	17.52941	5.40632	5
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D17. Answer opportunities analysis – DKC 17.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	8.39815	7.53551	8.72000	7.72598	3
#02 vanilla+forgets	7.29630	7.80637	7.80000	8.05890	2
#03 multigs T	3.32407	5.74022	3.44000	5.87307	1
#04 multigs+forgets T	3.00000	5.25072	2.98667	5.17659	1
#05 multilearn T	8.39815	7.53551	8.72000	7.72598	3
#06 multilearn+forgets T	9.06481	7.77039	9.30667	7.95252	3
#07 multigs+multilearn T	3.32407	5.74022	3.44000	5.87307	1
#08 multigs+multilearn+forgets T	1.71296	3.11178	1.69333	3.03125	1
#09 multiprior T	12.25000	6.41424	12.46667	6.49809	6
#10 multiprior+forgets T	11.90741	6.47286	12.18667	6.54649	6
#11 multigs TE	4.30556	6.69141	4.74667	7.13024	1
#12 multigs+forgets TE	3.75926	6.20050	4.01333	6.49219	1
#13 multilearn TE	8.39815	7.53551	8.72000	7.72598	3
#14 multilearn+forgets TE	7.97222	8.06684	8.70667	8.35815	2
#15 multigs+multilearn TE	4.30556	6.69141	4.74667	7.13024	1
#16 multigs+multilearn+forgets TE	4.77778	7.00512	5.42667	7.51606	1
#17 multiprior TE	12.25000	6.41424	12.46667	6.49809	6
#18 multiprior+forgets TE	11.90741	6.47286	12.18667	6.54649	6

Table D18. Answer opportunities analysis – DKC 19.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	12.72566	8.47396	12.70886	8.55323	2
#02 vanilla+forgets	20.00000	0.00000	20.00000	0.00000	20
#03 multigs T	10.35398	9.05826	10.50633	9.09053	1
#04 multigs+forgets T	20.00000	0.00000	20.00000	0.00000	20
#05 multilearn T	12.72566	8.47396	12.70886	8.55323	2
#06 multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#07 multigs+multilearn T	10.35398	9.05826	10.50633	9.09053	1
#08 multigs+multilearn+forgets T	20.00000	0.00000	20.00000	0.00000	20
#09 multiprior T	14.69912	7.14503	14.83544	7.09718	4
#10 multiprior+forgets T	20.00000	0.00000	20.00000	0.00000	20
#11 multigs TE	10.35398	9.05826	10.50633	9.09053	1
#12 multigs+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#13 multilearn TE	12.72566	8.47396	12.70886	8.55323	2
#14 multilearn+forgets TE	20.00000	0.00000	20.00000	0.00000	20
#15 multigs+multilearn TE	10.35398	9.05826	10.50633	9.09053	1
#16 multigs+multilearn+forgets TE	12.49558	8.78119	12.43038	8.89783	1
#17 multiprior TE	14.69912	7.14503	14.83544	7.09718	4
#18 multiprior+forgets TE	20.00000	0.00000	20.00000	0.00000	20

Table D19. Answer opportunities analysis – DKC 21.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	7.54386	8.73417	7.65000	8.94300	1
#02 vanilla+forgets	8.48246	8.26254	8.66250	8.48221	2
#03 multigs T	6.54386	8.36781	6.76250	8.57025	1
#04 multigs+forgets T	6.33333	8.25658	6.50000	8.42089	1
#05 multilearn T	7.55263	8.74424	7.65000	8.94300	1
#06 multilearn+forgets T	7.87719	8.67520	8.06250	8.89509	1
#07 multigs+multilearn T	6.54386	8.36781	6.76250	8.57025	1
#08 multigs+multilearn+forgets T	6.33333	8.25658	6.50000	8.42089	1
#09 multiprior T	10.77193	7.43828	10.76250	7.58453	4
#10 multiprior+forgets T	11.28947	7.55947	11.05000	7.63553	4
#11 multigs TE	6.54386	8.36781	6.76250	8.57025	1
#12 multigs+forgets TE	6.33333	8.25658	6.50000	8.42089	1
#13 multilearn TE	7.57018	8.76737	7.65000	8.94300	1
#14 multilearn+forgets TE	8.36842	8.98697	8.41250	9.12320	1
#15 multigs+multilearn TE	6.54386	8.36781	6.76250	8.57025	1
#16 multigs+multilearn+forgets TE	7.34211	8.59700	7.57500	8.79985	1
#17 multiprior TE	10.77193	7.43828	10.76250	7.58453	4
#18 multiprior+forgets TE	11.28947	7.55947	11.05000	7.63553	4

Table D20. Answer opportunities analysis – DKC 23.

BKT model	Average #AO (complete dataset)	SD	Average #AO (testing subset)	SD	Ideal learning path (#AO)
#01 vanilla	9.61458	8.24158	9.61194	8.40987	2
#02 vanilla+forgets	9.63542	8.25019	9.62687	8.42765	2
#03 multigs T	7.98958	8.44486	8.10448	8.62894	1
#04 multigs+forgets T	7.98958	8.44486	8.10448	8.62894	1
#05 multilearn T	9.61458	8.24158	9.61194	8.40987	2
#06 multilearn+forgets T	9.69792	8.22079	9.68657	8.40672	2
#07 multigs+multilearn T	7.98958	8.44486	8.10448	8.62894	1
#08 multigs+multilearn+forgets T	8.08333	8.52818	8.10448	8.62894	1
#09 multiprior T	10.89583	7.62886	11.11940	7.77822	3
#10 multiprior+forgets T	11.59375	7.15185	11.82090	7.32353	4
#11 multigs TE	7.98958	8.44486	8.10448	8.62894	1
#12 multigs+forgets TE	7.98958	8.44486	8.10448	8.62894	1
#13 multilearn TE	9.61458	8.24158	9.61194	8.40987	2
#14 multilearn+forgets TE	9.67708	8.23663	9.65672	8.42894	2
#15 multigs+multilearn TE	7.98958	8.44486	8.10448	8.62894	1
#16 multigs+multilearn+forgets TE	8.00000	8.45546	8.11940	8.64366	1
#17 multiprior TE	10.89583	7.62886	11.11940	7.77822	3
#18 multiprior+forgets TE	11.59375	7.15185	11.82090	7.32353	4

Table D21. Answer opportunities analysis – DKC 25.

APPENDIX E

BKT model ranking (normalised features, KM estimation based on the testing subset) - DKC 01-DKC 25

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.65640	0.59061	0.99630	0.42046	0.59861	0.92857	0.92325	1.00000	0.59456	0.89474	0.76035	6	0.77686	6
#02 vanilla+forgets	0.74270	0.73595	0.91986	0.72589	0.47207	0.94545	0.98333	0.96131	0.55102	0.89474	0.79323	3	0.79829	3
#03 multigs T	0.33356	0.60121	0.00808	0.02208	0.01134	0.75362	0.94226	0.99524	1.00000	1.00000	0.56674	13	0.61608	11
#04 multigs+forgets T	1.00000	0.93134	0.38244	0.90026	0.71395	0.00000	0.00000	0.00000	0.00000	0.00000	0.39280	14	0.38563	14
#05 multilearn T	0.64416	0.58794	0.97342	0.37105	0.71233	0.92857	0.92325	1.00000	0.62177	0.94737	0.77098	5	0.78387	5
#06 multilearn+forgets T	0.80790	0.70489	1.00000	0.83996	0.56699	1.00000	0.98804	0.95597	0.52109	0.89474	0.82796	1	0.82144	1
#07 multigs+multilearn T	0.33593	0.62204	0.02108	0.00000	0.02607	0.75362	0.94226	0.99524	1.00000	1.00000	0.56962	12	0.62211	10
#08 multigs+multilearn+forgets T	0.81484	0.82692	0.36138	1.00000	0.61885	0.00000	0.00000	0.00000	0.00000	0.00000	0.36220	15	0.33386	15
#09 multiprior T	0.00000	0.00000	0.96561	0.68523	0.44723	0.96296	1.00000	0.97879	0.49252	0.84211	0.63744	9	0.54387	13
#10 multiprior+forgets T	0.26537	0.17720	0.08661	0.78247	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.23117	18	0.08820	18
#11 multigs TE	0.45167	0.89392	0.00000	0.02776	0.00218	0.75362	0.94226	0.99524	1.00000	1.00000	0.60667	11	0.68320	9
#12 multigs+forgets TE	0.63811	1.00000	0.02294	0.09808	0.00000	0.75362	0.94226	0.99524	1.00000	1.00000	0.64503	7	0.73578	7
#13 multilearn TE	0.64536	0.59826	0.96938	0.37047	0.70757	0.92857	0.92325	1.00000	0.62177	0.94737	0.77120	4	0.78512	4
#14 multilearn+forgets TE	0.77638	0.74655	0.98875	0.83699	0.55803	1.00000	0.98804	0.95597	0.50340	0.89474	0.82489	2	0.81830	2
#15 multigs+multilearn TE	0.45154	0.91456	0.02349	0.00527	0.02789	0.75362	0.94226	0.99524	1.00000	1.00000	0.61139	10	0.69054	8
#16 multigs+multilearn+forgets TE	0.84454	0.89613	0.22332	0.99152	0.57367	0.00000	0.00000	0.00000	0.00000	0.00000	0.35292	16	0.32733	16
#17 multiprior TE	0.02218	0.01848	0.96561	0.68523	0.44723	0.96296	1.00000	0.97879	0.49252	0.84211	0.64151	8	0.55064	12
#18 multiprior+forgets TE	0.29145	0.19465	0.08661	0.78247	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.23552	17	0.09545	17

Table E1. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 01.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.62366	0.15843	0.97072	1.00000	1.00000	0.97531	1.00000	0.99510	0.53631	0.89474	0.81543	4	0.69319	10
#02 vanilla+forgets	1.00000	0.58815	0.94391	0.57302	0.80124	1.00000	0.94837	0.94203	0.51356	0.89474	0.82050	3	0.82339	2
#03 multigs T	0.29711	0.42684	0.26243	0.52521	0.23445	0.77451	0.90144	0.79868	0.93263	1.00000	0.61533	13	0.61559	14
#04 multigs+forgets T	0.30048	0.82783	0.00473	0.00269	0.03124	0.74528	0.89423	0.79484	1.00000	1.00000	0.56013	16	0.64639	12
#05 multilearn T	0.64099	0.16828	0.99557	0.85420	0.84906	0.97336	0.97508	0.97015	0.50831	0.89474	0.78297	6	0.69688	8
#06 multilearn+forgets T	0.95154	0.65214	1.00000	0.96082	0.85612	0.94530	0.96154	1.00000	0.40682	0.84211	0.85764	1	0.79965	3
#07 multigs+multilearn T	0.30139	0.42508	0.26412	0.52545	0.23613	0.77451	0.90144	0.79868	0.93263	1.00000	0.61594	12	0.61629	13
#08 multigs+multilearn+forgets T	0.44194	0.69826	0.14150	0.28165	0.00000	0.75962	0.88305	0.78168	0.94576	1.00000	0.59335	15	0.66451	11
#09 multiprior T	0.00000	0.00000	0.97347	0.69075	0.81197	0.98685	0.96154	0.95588	0.35258	0.73684	0.64699	10	0.50829	15.5
#10 multiprior+forgets T	0.41051	0.16760	0.74240	0.81745	0.92653	0.00000	0.00000	0.00000	0.00000	0.00000	0.30645	17	0.22008	17
#11 multigs TE	0.65965	0.85815	0.24359	0.53558	0.24103	0.77451	0.90144	0.79868	0.92738	1.00000	0.69400	7	0.74388	5
#12 multigs+forgets TE	0.67446	1.00000	0.00000	0.00000	0.02836	0.74528	0.89423	0.79484	1.00000	1.00000	0.61372	14	0.73662	7
#13 multilearn TE	0.63540	0.17221	0.98347	0.81074	0.88732	0.97436	0.98788	0.98299	0.51794	0.89474	0.78470	5	0.69635	9
#14 multilearn+forgets TE	0.96031	0.65607	0.98417	0.71118	0.83067	0.98685	0.96154	0.95588	0.47857	0.89474	0.84200	2	0.82678	1
#15 multigs+multilearn TE	0.62773	0.84102	0.25699	0.51871	0.22822	0.77451	0.90144	0.79868	0.93263	1.00000	0.68799	8	0.73881	6
#16 multigs+multilearn+forgets TE	0.88292	0.92782	0.16454	0.32273	0.01617	0.77451	0.90144	0.79868	0.94751	1.00000	0.67363	9	0.78288	4
#17 multiprior TE	0.00000	0.00000	0.97347	0.69075	0.81197	0.98685	0.96154	0.95588	0.35258	0.73684	0.64699	11	0.50829	15.5
#18 multiprior+forgets TE	0.41051	0.16760	0.74240	0.81745	0.92653	0.00000	0.00000	0.00000	0.00000	0.00000	0.30645	18	0.22008	18

Table E2. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 02.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.03818	0.00000	0.78237	0.90720	0.79471	0.71758	0.98611	0.98529	0.76714	0.94737	0.69259	5	0.54211	9
#02 vanilla+forgets	0.50878	0.55081	0.35787	0.18677	0.30996	0.00000	0.00000	0.00000	0.00000	0.00000	0.19142	15	0.23624	16
#03 multigs T	0.00125	0.48373	0.26279	0.47612	0.99994	0.63656	0.92869	0.96205	1.00000	1.00000	0.67511	7	0.56406	4
#04 multigs+forgets T	0.66129	0.82808	0.00541	0.00000	0.59927	0.00000	0.00000	0.00000	0.00000	0.00000	0.20941	13	0.24914	13
#05 multilearn T	0.05631	0.00167	0.81396	1.00000	0.88717	0.73086	1.00000	1.00000	0.75938	0.94737	0.71967	1	0.55159	7
#06 multilearn+forgets T	0.51819	0.55552	0.38310	0.19963	0.25825	0.00000	0.00000	0.00000	0.00000	0.00000	0.19147	14	0.24280	14
#07 multigs+multilearn T	0.00000	0.48339	0.26294	0.47418	1.00000	0.63656	0.92869	0.96205	1.00000	1.00000	0.67478	8	0.56381	5
#08 multigs+multilearn+forgets T	0.73630	0.84373	0.02140	0.02794	0.55841	0.00000	0.00000	0.00000	0.00000	0.00000	0.21879	12	0.26692	12
#09 multiprior T	0.21134	0.17711	1.00000	0.90157	0.84710	1.00000	0.78889	0.82462	0.26908	0.73684	0.67565	6	0.56573	3
#10 multiprior+forgets T	0.46221	0.43590	0.00000	0.00920	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.09073	17	0.14969	17
#11 multigs TE	0.31323	0.55547	0.29588	0.50605	0.92137	0.64699	0.94060	0.94023	0.99094	1.00000	0.71108	3	0.63375	1
#12 multigs+forgets TE	1.00000	1.00000	0.27170	0.16026	0.50440	0.00000	0.00000	0.00000	0.00000	0.00000	0.29364	10	0.37862	10
#13 multilearn TE	0.04776	0.00108	0.78271	0.90982	0.79135	0.71758	0.98611	0.98529	0.76714	0.94737	0.69362	4	0.54394	8
#14 multilearn+forgets TE	0.51767	0.57161	0.32821	0.17446	0.31308	0.00000	0.00000	0.00000	0.00000	0.00000	0.19050	16	0.23625	15
#15 multigs+multilearn TE	0.29460	0.55385	0.30101	0.51641	0.94577	0.64699	0.94060	0.94023	0.99483	1.00000	0.71343	2	0.63188	2
#16 multigs+multilearn+forgets TE	0.96757	0.99112	0.08714	0.06571	0.49849	0.00000	0.00000	0.00000	0.00000	0.00000	0.26100	11	0.34097	11
#17 multiprior TE	0.20191	0.13470	1.00000	0.90157	0.84710	1.00000	0.78889	0.82462	0.26908	0.73684	0.67047	9	0.55709	6
#18 multiprior+forgets TE	0.42019	0.41453	0.00000	0.00920	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.08439	18	0.13912	18

Table E3. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 03.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.26637	0.10001	0.81141	0.74957	0.91721	0.96364	0.96154	0.96154	0.79275	0.94737	0.74714	8	0.64692	14
#02 vanilla+forgets	0.49156	0.36054	0.77565	0.71841	0.81736	0.96364	0.96154	0.96154	0.79275	0.94737	0.77903	4	0.72192	5
#03 multigs T	0.53143	0.64215	0.00464	0.16442	0.19336	0.88333	0.92336	0.92593	0.98618	1.00000	0.62548	15	0.67462	8
#04 multigs+forgets T	0.97282	1.00000	0.34378	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.23166	18	0.38610	18
#05 multilearn T	0.29482	0.12504	0.89385	0.76670	1.00000	0.98148	0.98039	0.98039	0.78238	0.94737	0.77524	5	0.67082	11
#06 multilearn+forgets T	0.51095	0.20580	0.87444	0.76411	0.97395	0.98148	0.98039	0.98039	0.76684	0.94737	0.79857	2	0.71448	6
#07 multigs+multilearn T	0.51072	0.63669	0.00904	0.16123	0.17916	0.88333	0.92336	0.92593	0.99482	1.00000	0.62243	17	0.67243	10
#08 multigs+multilearn+forgets T	0.87973	0.90174	0.00365	0.15514	0.17878	0.88333	0.92336	0.92593	0.99482	1.00000	0.68465	13	0.77721	3
#09 multiprior T	0.00000	0.00000	0.91452	1.00000	0.86560	0.98148	0.98039	0.98039	0.70812	0.89474	0.73252	10	0.58314	17
#10 multiprior+forgets T	0.15398	0.15974	0.97843	0.99176	0.86267	1.00000	1.00000	1.00000	0.68394	0.89474	0.77253	6	0.64514	15
#11 multigs TE	0.61010	0.58116	0.00722	0.16068	0.18814	0.88333	0.92336	0.92593	0.99482	1.00000	0.62747	14	0.67944	7
#12 multigs+forgets TE	0.87509	0.94503	0.00504	0.16215	0.20283	0.88333	0.92336	0.92593	1.00000	1.00000	0.69228	12	0.78475	2
#13 multilearn TE	0.27043	0.11394	0.86368	0.75645	0.94786	0.98148	0.98039	0.98039	0.78929	0.94737	0.76313	7	0.66103	12
#14 multilearn+forgets TE	0.50268	0.36016	1.00000	0.99043	0.87554	1.00000	1.00000	1.00000	0.73921	0.94737	0.84154	1	0.75824	4
#15 multigs+multilearn TE	0.60194	0.56026	0.00000	0.15839	0.18426	0.88333	0.92336	0.92593	0.99482	1.00000	0.62323	16	0.67339	9
#16 multigs+multilearn+forgets TE	1.00000	0.89665	0.01579	0.15862	0.18697	0.88333	0.92336	0.92593	1.00000	1.00000	0.69906	11	0.79930	1
#17 multiprior TE	0.08331	0.01335	0.91452	1.00000	0.86560	0.98148	0.98039	0.98039	0.70812	0.89474	0.74219	9	0.59925	16
#18 multiprior+forgets TE	0.23308	0.16850	0.97843	0.99176	0.86267	1.00000	1.00000	1.00000	0.68394	0.89474	0.78131	3	0.65978	13

Table E4. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 04.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.06599	0.00000	0.97143	0.76719	1.00000	0.42667	0.94251	0.99821	0.72254	0.94444	0.68390	5	0.52185	8
#02 vanilla+forgets	0.74782	0.70483	0.55934	0.97505	0.12293	0.00000	0.00000	0.00000	0.00000	0.00000	0.31100	13	0.33533	14
#03 multigs T	0.00000	0.05032	0.89434	0.65828	0.58797	0.38095	1.00000	1.00000	0.99422	1.00000	0.65661	7	0.55331	5
#04 multigs+forgets T	0.76902	0.76070	0.01871	0.92316	0.00905	0.00000	0.00000	0.00000	0.00000	0.00000	0.24806	17	0.25807	17
#05 multilearn T	0.06557	0.00277	0.97226	0.76822	0.99812	0.42667	0.94251	0.99821	0.72254	0.94444	0.68413	4	0.52238	7
#06 multilearn+forgets T	0.74958	0.70285	0.74626	0.00000	0.10281	1.00000	0.30719	0.40062	0.19075	1.00000	0.52001	10	0.73157	1
#07 multigs+multilearn T	0.00760	0.06933	0.89273	0.64015	0.59409	0.38095	1.00000	1.00000	1.00000	1.00000	0.65849	6	0.55844	4
#08 multigs+multilearn+forgets T	0.76933	0.76426	0.00000	0.92420	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.24578	18	0.25560	18
#09 multiprior T	0.17712	0.00137	1.00000	0.14523	0.71316	0.48485	0.80918	0.91313	0.56358	0.88889	0.56965	8.5	0.51930	10
#10 multiprior+forgets T	0.60556	0.33636	0.82228	1.00000	0.04449	0.00000	0.00000	0.00000	0.00000	0.00000	0.28087	15	0.29403	15
#11 multigs TE	0.27532	0.44414	0.90695	0.72645	0.56181	0.38095	1.00000	1.00000	0.89017	1.00000	0.71858	2	0.64959	3
#12 multigs+forgets TE	0.96822	0.99921	0.12785	0.93368	0.03964	0.00000	0.00000	0.00000	0.00000	0.00000	0.30686	14	0.34921	12
#13 multilearn TE	0.06593	0.00812	0.97367	0.76998	0.99575	0.42667	0.94251	0.99821	0.72254	0.94444	0.68478	3	0.52356	6
#14 multilearn+forgets TE	0.74973	0.69315	0.65200	0.99372	0.10600	0.00000	0.00000	0.00000	0.00000	0.00000	0.31946	12	0.34915	13
#15 multigs+multilearn TE	0.29497	0.46672	0.90958	0.72340	0.56370	0.38095	1.00000	1.00000	0.89017	1.00000	0.72295	1	0.65707	2
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.48016	0.96074	0.09808	0.00000	0.00000	0.00000	0.00000	0.00000	0.35390	11	0.41336	11
#17 multiprior TE	0.17712	0.00137	1.00000	0.14523	0.71316	0.48485	0.80918	0.91313	0.56358	0.88889	0.56965	8.5	0.51930	9
#18 multiprior+forgets TE	0.60556	0.33636	0.82228	1.00000	0.04449	0.00000	0.00000	0.00000	0.00000	0.00000	0.28087	16	0.29403	16

Table E5. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 05.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.08740	0.45287	0.67871	1.00000	0.05098	0.65185	1.00000	0.65152	0.76344	1.00000	0.63368	8	0.60571	11
#02 vanilla+forgets	0.01532	0.77487	0.46455	0.67631	0.03476	0.65185	1.00000	0.65152	0.76344	1.00000	0.60326	12	0.61167	10
#03 multigs T	0.40218	0.77638	0.61806	0.81601	0.02265	0.65185	1.00000	0.65152	0.77419	1.00000	0.67128	3	0.70378	5
#04 multigs+forgets T	0.37951	0.87930	0.06597	0.22876	0.05775	0.00000	0.11312	0.00000	1.00000	1.00000	0.37248	17	0.55419	13
#05 multilearn T	0.14170	0.50433	0.71211	0.75605	0.05706	0.65185	1.00000	0.65152	0.72043	1.00000	0.61951	9	0.62174	8
#06 multilearn+forgets T	0.12909	0.59893	0.65748	0.93731	0.06362	0.65185	1.00000	0.65152	0.75269	1.00000	0.64425	7	0.63167	7
#07 multigs+multilearn T	0.43820	0.72606	0.57766	0.79753	0.00000	0.65185	1.00000	0.65152	0.76344	1.00000	0.66063	6	0.69287	6
#08 multigs+multilearn+forgets T	0.26665	0.92470	0.00000	0.11850	0.03138	0.00000	0.11312	0.00000	1.00000	1.00000	0.34543	18	0.53189	14
#09 multiprior T	0.00000	0.00000	1.00000	0.21345	1.00000	1.00000	0.00000	1.00000	0.00000	0.00000	0.42135	15	0.33333	17
#10 multiprior+forgets T	0.20258	0.30678	0.94497	0.00000	0.98160	1.00000	0.00000	1.00000	0.11828	0.00000	0.45542	13	0.42877	15.5
#11 multigs TE	1.00000	0.96708	0.62755	0.61559	0.01460	0.65185	1.00000	0.65152	0.73118	1.00000	0.72594	2	0.82961	1
#12 multigs+forgets TE	0.79481	1.00000	0.30490	0.47560	0.03253	0.65185	1.00000	0.65152	0.77419	1.00000	0.66854	5	0.75429	3
#13 multilearn TE	0.08801	0.45552	0.68600	0.75314	0.06049	0.65185	1.00000	0.65152	0.73118	1.00000	0.60777	11	0.60209	12
#14 multilearn+forgets TE	0.03216	0.78017	0.49032	0.73402	0.04397	0.65185	1.00000	0.65152	0.76344	1.00000	0.61474	10	0.61966	9
#15 multigs+multilearn TE	0.96973	0.95043	0.58563	0.85667	0.02264	0.65185	1.00000	0.65152	0.76344	1.00000	0.74519	1	0.82018	2
#16 multigs+multilearn+forgets TE	0.78285	0.94627	0.24827	0.50556	0.09484	0.65185	1.00000	0.65152	0.81720	1.00000	0.66984	4	0.74108	4
#17 multiprior TE	0.00000	0.00000	1.00000	0.21345	1.00000	1.00000	0.00000	1.00000	0.00000	0.00000	0.42135	16	0.33333	18
#18 multiprior+forgets TE	0.20258	0.30678	0.94497	0.00000	0.98160	1.00000	0.00000	1.00000	0.11828	0.00000	0.45542	14	0.42877	15.5

Table E6. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 06.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.13341	0.96209	0.21430	0.18926	0.69875	0.45000	0.51923	0.51852	0.96970	1.00000	0.56553	9	0.62158	8
#02 vanilla+forgets	0.71863	0.80918	0.28915	0.08579	0.38705	0.45000	0.51923	0.51852	0.84848	1.00000	0.56260	10	0.68591	1
#03 multigs T	0.00000	0.96595	0.01286	0.13311	0.99228	0.00000	1.00000	1.00000	1.00000	1.00000	0.61042	5	0.49647	14
#04 multigs+forgets T	0.63602	0.88114	0.08760	0.09514	0.90647	0.45000	0.51923	0.51852	0.96970	1.00000	0.60638	6	0.67074	2
#05 multilearn T	0.20504	1.00000	0.16223	0.20581	0.80242	0.45000	0.51923	0.51852	0.96970	1.00000	0.58329	7	0.63116	6
#06 multilearn+forgets T	0.52542	0.84773	0.32065	0.01070	0.19185	0.45000	0.51923	0.51852	0.30303	0.00000	0.36871	17	0.40780	15
#07 multigs+multilearn T	0.00909	0.96595	0.01200	0.12845	0.99363	0.00000	1.00000	1.00000	1.00000	1.00000	0.61091	4	0.49784	13
#08 multigs+multilearn+forgets T	0.60600	0.88885	0.01242	0.07463	0.98782	0.00000	1.00000	1.00000	0.45455	0.00000	0.50243	14	0.32697	18
#09 multiprior T	0.48934	0.00000	0.82700	1.00000	0.76990	1.00000	0.00000	0.00000	0.00000	0.94444	0.50307	12.5	0.54347	11.5
#10 multiprior+forgets T	0.90788	0.00438	1.00000	0.80679	0.00000	1.00000	0.00000	0.00000	0.00000	0.94444	0.46635	16	0.64278	4.5
#11 multigs TE	0.29670	0.96595	0.00661	0.11903	1.00000	0.00000	1.00000	1.00000	1.00000	1.00000	0.63883	3	0.54488	10
#12 multigs+forgets TE	1.00000	0.88371	0.00018	0.10713	0.99012	0.00000	1.00000	1.00000	1.00000	1.00000	0.69811	1	0.64731	3
#13 multilearn TE	0.20477	0.99936	0.16228	0.20566	0.80242	0.45000	0.51923	0.51852	0.96970	1.00000	0.58319	8	0.63102	7
#14 multilearn+forgets TE	0.52738	0.81304	0.32365	0.00000	0.17977	0.45000	0.51923	0.51852	0.30303	0.00000	0.36346	18	0.40285	16
#15 multigs+multilearn TE	0.33819	0.96980	0.00000	0.13698	0.99764	0.00000	1.00000	1.00000	1.00000	1.00000	0.64426	2	0.55133	9
#16 multigs+multilearn+forgets TE	0.94842	0.88628	0.00215	0.07951	0.99407	0.00000	1.00000	1.00000	0.45455	0.00000	0.53650	11	0.38190	17
#17 multiprior TE	0.48934	0.00000	0.82700	1.00000	0.76990	1.00000	0.00000	0.00000	0.00000	0.94444	0.50307	12.5	0.54347	11.5
#18 multiprior+forgets TE	0.90788	0.00438	1.00000	0.80679	0.00000	1.00000	0.00000	0.00000	0.00000	0.94444	0.46635	15	0.64278	4.5

Table E7. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 07.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.27566	0.25515	0.99546	0.97099	0.87300	0.99040	0.99057	0.99020	0.46659	0.89474	0.77027	2	0.64633	8
#02 vanilla+forgets	0.26248	0.17639	0.98429	0.94475	0.81292	1.00000	0.98131	0.98058	0.44332	0.84211	0.74281	5	0.61810	12
#03 multigs T	0.99279	0.99886	0.31206	0.27896	0.03694	0.89166	0.87979	0.86325	0.91229	1.00000	0.71666	10	0.85128	2
#04 multigs+forgets T	0.79334	0.94273	0.00000	0.00000	0.24925	0.86271	0.84859	0.85065	1.00000	1.00000	0.65473	17	0.76646	7
#05 multilearn T	0.24391	0.23839	0.99006	0.97149	0.85989	0.99040	0.99057	0.99020	0.46122	0.89474	0.76308	3	0.63645	10
#06 multilearn+forgets T	0.25589	0.17744	0.98492	0.94857	0.81646	1.00000	0.98131	0.98058	0.43496	0.84211	0.74222	6	0.61589	13
#07 multigs+multilearn T	1.00000	0.99664	0.35009	0.27822	0.00000	0.89768	0.88630	0.86984	0.90274	1.00000	0.71815	9	0.85786	1
#08 multigs+multilearn+forgets T	0.92466	0.96630	0.03948	0.03267	0.28500	0.86271	0.84859	0.85065	0.99284	1.00000	0.68029	16	0.79767	6
#09 multiprior T	0.00380	0.04725	0.99257	0.96585	1.00000	0.99040	0.99057	0.99020	0.36038	0.73684	0.70778	11	0.52187	14
#10 multiprior+forgets T	0.03489	0.00545	0.99968	0.92783	0.94872	0.96894	0.95759	0.97776	0.33890	0.73684	0.68966	13	0.51412	16
#11 multigs TE	0.96510	1.00000	0.31257	0.30702	0.08314	0.89166	0.87979	0.86325	0.92780	1.00000	0.72303	7	0.84952	4
#12 multigs+forgets TE	0.80183	0.98215	0.84765	0.67331	0.29852	0.00000	0.00000	0.00000	0.00000	0.00000	0.36035	18	0.43861	18
#13 multilearn TE	0.24609	0.26314	1.00000	1.00000	0.87050	0.99879	1.00000	1.00000	0.45764	0.89474	0.77309	1	0.64340	9
#14 multilearn+forgets TE	0.26222	0.19865	0.99330	0.95624	0.82702	0.99040	0.99057	0.99020	0.45406	0.89474	0.75574	4	0.63223	11
#15 multigs+multilearn TE	0.97202	0.99151	0.31870	0.30251	0.04719	0.89166	0.87979	0.86325	0.92780	1.00000	0.71944	8	0.85028	3
#16 multigs+multilearn+forgets TE	0.96067	0.96734	0.02897	0.03720	0.30623	0.86271	0.84859	0.85065	0.99284	1.00000	0.68552	15	0.80209	5
#17 multiprior TE	0.00000	0.03826	0.99257	0.96585	1.00000	0.99040	0.99057	0.99020	0.36038	0.73684	0.70651	12	0.51974	15
#18 multiprior+forgets TE	0.03150	0.00000	0.99968	0.92783	0.94872	0.96894	0.95759	0.97776	0.33890	0.73684	0.68878	14	0.51265	17

Table E8. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 08.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.16407	0.06435	1.00000	0.99887	0.99676	0.67442	1.00000	1.00000	0.77994	0.94444	0.76229	5	0.60454	6
#02 vanilla+forgets	0.62010	0.67343	0.53599	0.38843	0.44576	0.00000	0.00000	0.00000	0.00000	0.00000	0.26637	12	0.30492	13
#03 multigs T	0.28485	0.39835	0.82552	0.71333	0.56930	0.59184	0.99781	0.95782	1.00000	1.00000	0.73388	7	0.68343	4
#04 multigs+forgets T	0.59969	0.78079	0.16068	0.31272	0.34160	0.00000	0.00000	0.00000	0.00000	0.00000	0.21955	15	0.25686	16
#05 multilearn T	0.16473	0.06322	0.99925	1.00000	1.00000	0.67442	1.00000	1.00000	0.77994	0.94444	0.76260	4	0.60433	7
#06 multilearn+forgets T	0.61881	0.66610	0.47557	0.33150	0.38548	0.00000	0.00000	0.00000	0.00000	0.00000	0.24775	13	0.29341	14
#07 multigs+multilearn T	0.28481	0.39858	0.82571	0.71350	0.56976	0.59184	0.99781	0.95782	1.00000	1.00000	0.73398	6	0.68349	3
#08 multigs+multilearn+forgets T	0.78972	0.86924	0.68770	0.68272	0.64544	0.00000	0.00000	0.00000	0.00000	0.00000	0.36748	11	0.39111	11
#09 multiprior T	0.00000	0.00000	0.98139	0.61498	0.66252	1.00000	0.72917	0.73143	0.34304	0.77778	0.58403	9	0.51703	9
#10 multiprior+forgets T	0.53720	0.45579	0.27148	0.11856	0.17719	0.00000	0.00000	0.00000	0.00000	0.00000	0.15602	18	0.21074	18
#11 multigs TE	0.44739	0.63170	0.91735	0.74434	0.71692	0.63043	0.95890	0.95522	0.91748	1.00000	0.79197	1	0.75739	2
#12 multigs+forgets TE	1.00000	1.00000	0.89557	0.80468	0.73112	0.00000	0.00000	0.00000	0.00000	0.00000	0.44314	10	0.48259	10
#13 multilearn TE	0.16477	0.06461	0.99920	0.99996	0.99991	0.67442	1.00000	1.00000	0.77994	0.94444	0.76272	3	0.60456	5
#14 multilearn+forgets TE	0.62589	0.66253	0.41740	0.24903	0.27403	0.00000	0.00000	0.00000	0.00000	0.00000	0.22289	14	0.28430	15
#15 multigs+multilearn TE	0.44774	0.63187	0.91716	0.74425	0.71646	0.63043	0.95890	0.95522	0.91748	1.00000	0.79195	2	0.75745	1
#16 multigs+multilearn+forgets TE	0.93037	0.97245	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.19028	16	0.31714	12
#17 multiprior TE	0.01576	0.02877	0.98139	0.61498	0.66252	1.00000	0.72917	0.73143	0.34304	0.77778	0.58848	8	0.52446	8
#18 multiprior+forgets TE	0.57837	0.48209	0.27148	0.11856	0.17719	0.00000	0.00000	0.00000	0.00000	0.00000	0.16277	17	0.22199	17

Table E9. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 09.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.22202	0.02201	0.76951	0.36738	0.62568	0.59289	0.95652	0.95413	0.87841	0.94737	0.63359	9	0.57203	7
#02 vanilla+forgets	0.31495	0.27023	0.37182	0.06951	0.00249	0.00000	0.00000	0.00000	0.00000	0.00000	0.10290	15	0.15950	15
#03 multigs T	0.35494	0.28028	0.52471	0.31333	0.99998	0.56818	0.94638	0.99498	0.99922	1.00000	0.69820	4	0.62122	3
#04 multigs+forgets T	0.46290	0.51529	0.38427	0.04858	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.14110	14	0.22708	14
#05 multilearn T	0.22223	0.02151	0.77101	0.36760	0.62549	0.59289	0.95652	0.95413	0.87841	0.94737	0.63372	8	0.57223	6
#06 multilearn+forgets T	0.33097	0.26586	0.77488	0.58161	0.68615	0.62696	1.00000	1.00000	0.79735	0.00000	0.60638	11	0.46600	11
#07 multigs+multilearn T	0.34749	0.28010	0.52844	0.32011	1.00000	0.56818	0.94638	0.99498	1.00000	1.00000	0.69857	3	0.62070	4
#08 multigs+multilearn+forgets T	0.47104	0.55972	0.77359	0.39528	0.74083	0.63425	0.98515	0.95513	0.77241	0.00000	0.62874	10	0.53517	8
#09 multiprior T	0.00188	0.00000	1.00000	1.00000	0.88042	1.00000	0.78930	0.79710	0.19797	0.68421	0.63509	6	0.48068	10
#10 multiprior+forgets T	0.18619	0.08674	0.13989	0.00000	0.16602	0.00000	0.00000	0.00000	0.00000	0.00000	0.05788	18	0.06880	18
#11 multigs TE	0.86739	0.90089	0.61044	0.30696	0.80169	0.58027	0.94017	0.98758	0.97506	1.00000	0.79705	2	0.82234	2
#12 multigs+forgets TE	1.00000	1.00000	0.45718	0.09372	0.01100	0.00000	0.00000	0.00000	0.00000	0.00000	0.25619	12	0.40953	12
#13 multilearn TE	0.22226	0.02348	0.77111	0.36761	0.62544	0.59289	0.95652	0.95413	0.87841	0.94737	0.63392	7	0.57259	5
#14 multilearn+forgets TE	0.30601	0.28097	0.34034	0.06451	0.02114	0.00000	0.00000	0.00000	0.00000	0.00000	0.10130	16	0.15455	16
#15 multigs+multilearn TE	0.86738	0.90105	0.61051	0.30703	0.80159	0.58027	0.94017	0.98758	0.97506	1.00000	0.79706	1	0.82238	1
#16 multigs+multilearn+forgets TE	0.89650	0.93339	0.00000	0.03183	0.20353	0.00000	0.00000	0.00000	0.00000	0.00000	0.20652	13	0.30498	13
#17 multiprior TE	0.00000	0.02251	1.00000	1.00000	0.88042	1.00000	0.78930	0.79710	0.19797	0.68421	0.63715	5	0.48412	9
#18 multiprior+forgets TE	0.19358	0.10630	0.13989	0.00000	0.16602	0.00000	0.00000	0.00000	0.00000	0.00000	0.06058	17	0.07330	17

Table E10. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 10.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.34403	0.23067	0.86392	0.57881	0.90768	0.83906	0.16091	0.30693	0.46814	0.75000	0.54502	8	0.58264	7
#02 vanilla+forgets	0.33099	0.33236	0.89554	0.68259	0.89466	0.82958	0.35870	0.46970	0.31717	0.50000	0.56113	7	0.53427	10
#03 multigs T	0.21807	0.16604	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	1.00000	0.23841	18	0.39735	14
#04 multigs+forgets T	0.50314	0.67908	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	1.00000	1.00000	0.31822	16	0.53037	11
#05 multilearn T	0.36136	0.21992	0.92092	0.56183	0.93834	0.82958	0.35870	0.46970	0.43629	0.75000	0.58466	4	0.58634	6
#06 multilearn+forgets T	0.32278	0.32259	0.96533	0.81751	0.93606	0.91412	0.46044	0.55357	0.24792	0.50000	0.60403	2	0.54546	9
#07 multigs+multilearn T	0.41703	0.54686	0.03838	0.11187	0.03093	0.00000	0.00000	0.00000	0.98753	1.00000	0.31326	17	0.49830	13
#08 multigs+multilearn+forgets T	0.62046	0.54557	0.00086	0.00056	0.02854	0.00000	0.00000	0.00000	1.00000	1.00000	0.31960	15	0.52782	12
#09 multiprior T	0.00000	0.00000	0.84373	0.56846	0.84788	0.74633	0.25887	0.38750	0.22022	0.25000	0.41230	11	0.34338	16
#10 multiprior+forgets T	0.01498	0.06347	0.96209	1.00000	0.91418	0.79929	1.00000	1.00000	0.00000	0.00000	0.57540	6	0.30664	18
#11 multigs TE	0.90807	0.85205	0.07704	0.16118	0.15246	0.00000	0.00000	0.00000	0.95291	1.00000	0.41037	13	0.63168	4
#12 multigs+forgets TE	0.92344	1.00000	0.06435	0.04498	0.04235	0.00000	0.00000	0.00000	0.97784	1.00000	0.40530	14	0.66094	2
#13 multilearn TE	0.36533	0.23491	0.90816	0.54911	0.91357	0.82958	0.35870	0.46970	0.44044	0.75000	0.58195	5	0.58807	5
#14 multilearn+forgets TE	0.34921	0.37728	1.00000	0.80748	1.00000	1.00000	0.56416	0.63918	0.24931	0.50000	0.64866	1	0.57930	8
#15 multigs+multilearn TE	0.91040	0.85346	0.07745	0.16319	0.15695	0.00000	0.00000	0.00000	0.95291	1.00000	0.41144	12	0.63237	3
#16 multigs+multilearn+forgets TE	1.00000	0.98335	0.04200	0.06595	0.06285	0.00000	0.00000	0.00000	0.97922	1.00000	0.41334	10	0.66743	1
#17 multiprior TE	0.06915	0.04029	0.84373	0.56846	0.84788	0.74633	0.25887	0.38750	0.22022	0.25000	0.42324	9	0.36162	15
#18 multiprior+forgets TE	0.08559	0.11207	0.96209	1.00000	0.91418	0.79929	1.00000	1.00000	0.00000	0.00000	0.58727	3	0.32643	17

Table E11. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 11.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.12425	0.33289	0.80314	0.77333	0.17499	0.93939	1.00000	0.97619	0.74945	0.94737	0.68210	5	0.64942	9
#02 vanilla+forgets	0.36820	0.64111	0.57686	0.42358	0.88192	0.00000	0.00000	0.00000	0.00000	0.00000	0.28917	16	0.26436	16
#03 multigs T	0.14824	0.39008	0.16944	0.34435	0.52844	0.83784	0.96549	0.98839	0.96044	1.00000	0.63327	11	0.58434	10
#04 multigs+forgets T	0.48866	0.71663	0.00170	0.00000	0.04347	0.80519	0.96122	0.98795	1.00000	1.00000	0.60048	13	0.66870	6
#05 multilearn T	0.12937	0.33448	0.80039	0.76848	0.16596	0.93939	1.00000	0.97619	0.74945	0.94737	0.68111	7	0.65007	7
#06 multilearn+forgets T	0.35851	0.56651	0.56954	0.81470	0.19059	0.90842	0.97290	0.93375	0.61758	0.00000	0.59325	14	0.50343	14
#07 multigs+multilearn T	0.13519	0.39570	0.10640	0.18529	0.51837	0.82667	0.95439	0.97619	0.97912	1.00000	0.60773	12	0.57385	12
#08 multigs+multilearn+forgets T	0.49962	0.70765	0.08460	0.19730	0.48808	0.81579	0.97214	1.00000	0.98681	1.00000	0.67520	9	0.68241	5
#09 multiprior T	0.00000	0.00000	1.00000	1.00000	1.00000	1.00000	0.98605	0.99055	0.56813	0.84211	0.73868	2	0.56837	13
#10 multiprior+forgets T	0.30581	0.25927	0.52504	0.38482	0.81192	0.00000	0.00000	0.00000	0.00000	0.00000	0.22869	18	0.18169	18
#11 multigs TE	0.61178	0.79409	0.10424	0.18417	0.53305	0.82667	0.95439	0.97619	0.98022	1.00000	0.69648	3	0.71950	3
#12 multigs+forgets TE	0.98538	1.00000	0.00025	0.00110	0.02994	0.80519	0.96122	0.98795	1.00000	1.00000	0.67710	8	0.79847	2
#13 multilearn TE	0.12908	0.33280	0.80141	0.77004	0.16865	0.93939	1.00000	0.97619	0.74945	0.94737	0.68144	6	0.64992	8
#14 multilearn+forgets TE	0.36482	0.62692	0.65210	0.45432	0.33958	0.79487	0.67097	0.70503	0.36044	0.00000	0.49690	15	0.46652	15
#15 multigs+multilearn TE	0.60960	0.79369	0.09773	0.18463	0.53634	0.82667	0.95439	0.97619	0.98022	1.00000	0.69595	4	0.71798	4
#16 multigs+multilearn+forgets TE	1.00000	0.99088	0.00000	0.00305	0.00000	0.80519	0.96122	0.98795	1.00000	1.00000	0.67483	10	0.79935	1
#17 multiprior TE	0.04099	0.01225	1.00000	1.00000	1.00000	1.00000	0.98605	0.99055	0.56813	0.84211	0.74401	1	0.57725	11
#18 multiprior+forgets TE	0.36143	0.28548	0.52504	0.38482	0.81192	0.00000	0.00000	0.00000	0.00000	0.00000	0.23687	17	0.19533	17

Table E12. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 12.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.22800	0.29604	0.84956	0.58044	0.59443	0.93913	0.94236	0.97851	0.46914	0.89474	0.67723	6	0.61277	9
#02 vanilla+forgets	0.44037	0.44632	0.83349	0.98094	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.37011	14	0.28670	15
#03 multigs T	0.10950	0.37683	0.32333	0.54228	0.37218	0.71816	0.99402	1.00000	0.90035	1.00000	0.63367	12	0.57136	11
#04 multigs+forgets T	0.42141	0.71870	0.12057	0.26644	0.42679	0.68588	0.98333	0.98999	0.96120	1.00000	0.65743	8	0.65129	6
#05 multilearn T	0.23303	0.31354	0.84751	0.57916	0.59011	0.93913	0.94236	0.97851	0.46825	0.89474	0.67863	4	0.61603	7
#06 multilearn+forgets T	0.45810	0.41720	0.84912	0.97861	0.98741	0.00000	0.00000	0.00000	0.00000	0.00000	0.36904	15	0.28740	14
#07 multigs+multilearn T	0.11389	0.38148	0.32118	0.52743	0.36980	0.71816	0.99402	1.00000	0.90035	1.00000	0.63263	13	0.57251	10
#08 multigs+multilearn+forgets T	0.43940	0.67518	0.17939	0.19481	0.41898	0.69368	0.96263	0.96739	0.94797	1.00000	0.64794	10	0.65594	5
#09 multiprior T	0.00000	0.00000	1.00000	1.00000	0.57933	1.00000	1.00000	0.93165	0.26102	0.73684	0.65088	9	0.49964	13
#10 multiprior+forgets T	0.24207	0.08304	0.67219	0.89981	0.94997	0.00000	0.00000	0.00000	0.00000	0.00000	0.28471	18	0.16622	18
#11 multigs TE	0.47168	0.97835	0.32983	0.53187	0.38643	0.71816	0.99402	1.00000	0.90388	1.00000	0.73142	2	0.73365	4
#12 multigs+forgets TE	0.87466	0.98704	0.00000	0.00000	0.00000	0.67081	0.96327	0.96915	1.00000	1.00000	0.64649	11	0.75542	2
#13 multilearn TE	0.23205	0.30772	0.84913	0.57529	0.59141	0.93913	0.94236	0.97851	0.46914	0.89474	0.67795	5	0.61532	8
#14 multilearn+forgets TE	0.43140	0.44771	0.77569	0.94713	0.96046	0.00000	0.00000	0.00000	0.00000	0.00000	0.35624	16	0.27580	16
#15 multigs+multilearn TE	0.51393	0.99518	0.31974	0.49735	0.37563	0.71816	0.99402	1.00000	0.90388	1.00000	0.73179	1	0.74182	3
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.15922	0.19813	0.40246	0.72671	0.94011	0.97649	0.85714	1.00000	0.72603	3	0.79051	1
#17 multiprior TE	0.04329	0.03405	1.00000	1.00000	0.57933	1.00000	1.00000	0.93165	0.26102	0.73684	0.65862	7	0.51253	12
#18 multiprior+forgets TE	0.31324	0.13190	0.67219	0.89981	0.94997	0.00000	0.00000	0.00000	0.00000	0.00000	0.29671	17	0.18622	17

Table E13. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 13.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.22814	0.31284	0.94366	0.93274	0.49299	1.00000	0.97634	0.97101	0.77548	1.00000	0.76332	6	0.71002	4
#02 vanilla+forgets	0.48258	0.55941	0.67381	0.59667	0.98003	0.00000	0.00000	0.00000	0.00000	0.00000	0.32925	13	0.28597	14
#03 multigs T	0.16846	0.27773	0.92474	1.00000	0.54563	0.96429	0.98438	0.98529	0.78375	1.00000	0.76343	5	0.68649	8
#04 multigs+forgets T	0.46128	0.54376	0.64357	0.57789	0.99714	0.00000	0.00000	0.00000	0.00000	0.00000	0.32236	15	0.27477	16
#05 multilearn T	0.22834	0.29639	0.94255	0.91400	0.50440	1.00000	0.97634	0.97101	0.77548	1.00000	0.76085	8	0.70713	6
#06 multilearn+forgets T	0.48445	0.56484	0.69234	0.60629	0.97747	0.00000	0.00000	0.00000	0.00000	0.00000	0.33254	12	0.29027	13
#07 multigs+multilearn T	0.20459	0.29978	0.92482	0.97933	0.52364	0.98182	0.99824	1.00000	0.77824	1.00000	0.76905	4	0.69821	7
#08 multigs+multilearn+forgets T	0.46298	0.54959	0.67257	0.59335	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.32785	14	0.28086	15
#09 multiprior T	0.00000	0.00000	1.00000	0.90754	0.29086	1.00000	0.94018	0.97101	0.61846	0.88889	0.66169	10	0.58456	10
#10 multiprior+forgets T	0.38978	0.29588	0.35387	0.37132	0.85272	0.00000	0.00000	0.00000	0.00000	0.00000	0.22636	18	0.17326	18
#11 multigs TE	0.68956	0.79456	0.56010	0.79518	0.23356	0.83077	1.00000	0.98363	1.00000	1.00000	0.78874	2	0.81250	3
#12 multigs+forgets TE	0.86282	0.99569	0.00000	0.00000	0.44492	0.00000	0.00000	0.00000	0.00000	0.00000	0.23034	17	0.30975	11
#13 multilearn TE	0.22653	0.29950	0.94169	0.91285	0.50684	1.00000	0.97634	0.97101	0.77548	1.00000	0.76102	7	0.70720	5
#14 multilearn+forgets TE	0.49672	0.58513	0.73275	0.62749	0.97551	0.00000	0.00000	0.00000	0.00000	0.00000	0.34176	11	0.30243	12
#15 multigs+multilearn TE	0.69191	0.79564	0.56921	0.79856	0.21381	0.83077	1.00000	0.98363	1.00000	1.00000	0.78835	3	0.81459	2
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.60789	0.80285	0.00000	0.93103	0.99198	0.95714	0.83196	1.00000	0.81229	1	0.89515	1
#17 multiprior TE	0.03387	0.04789	1.00000	0.90754	0.29086	1.00000	0.94018	0.97101	0.61846	0.88889	0.66987	9	0.59818	9
#18 multiprior+forgets TE	0.43282	0.33130	0.35387	0.37132	0.85272	0.00000	0.00000	0.00000	0.00000	0.00000	0.23420	16	0.18633	17

Table E14. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 14.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.18783	0.08426	0.90901	0.67730	0.46274	0.96386	0.99754	0.95726	0.66696	0.94737	0.68541	3	0.62655	12
#02 vanilla+forgets	0.47554	0.47999	0.59667	0.88040	0.97079	0.00000	0.00000	0.00000	0.00000	0.00000	0.34034	16	0.25870	16
#03 multigs T	0.38932	0.56079	0.28362	0.18629	0.00000	0.83333	0.95023	0.95726	0.95079	1.00000	0.61116	14	0.66964	8
#04 multigs+forgets T	0.56285	0.79251	0.00000	0.00000	0.03294	0.80000	0.94025	0.94971	1.00000	1.00000	0.60773	15	0.69239	6
#05 multilearn T	0.19815	0.09497	0.90723	0.65651	0.44107	0.96386	0.99754	0.95726	0.66169	0.94737	0.68256	4	0.62888	11
#06 multilearn+forgets T	0.47157	0.45890	0.91826	0.95608	0.78790	0.84211	0.91498	0.83761	0.24517	0.00000	0.64326	11	0.48934	15
#07 multigs+multilearn T	0.38711	0.60261	0.28427	0.18653	0.00054	0.83333	0.95023	0.95726	0.95079	1.00000	0.61527	13	0.67635	7
#08 multigs+multilearn+forgets T	0.59876	0.79146	0.08769	0.09577	0.06116	0.81633	0.93103	0.93668	0.98155	1.00000	0.63004	12	0.71263	5
#09 multiprior T	0.00000	0.00000	1.00000	0.78363	0.60840	1.00000	1.00000	0.99556	0.54218	0.84211	0.67719	7	0.56405	14
#10 multiprior+forgets T	0.35597	0.40452	0.48855	0.81915	0.93068	0.00000	0.00000	0.00000	0.00000	0.00000	0.29989	18	0.20817	18
#11 multigs TE	0.74685	0.79405	0.28356	0.18636	0.00005	0.83333	0.95023	0.95726	0.95079	1.00000	0.67025	8	0.76810	3
#12 multigs+forgets TE	0.94730	1.00000	0.00819	0.00660	0.04018	0.80000	0.94025	0.94971	1.00000	1.00000	0.66922	10	0.79258	2
#13 multilearn TE	0.19820	0.10162	0.90976	0.65007	0.43451	0.96386	0.99754	0.95726	0.66169	0.94737	0.68219	5	0.63042	10
#14 multilearn+forgets TE	0.46338	0.46853	0.97496	1.00000	1.00000	0.93677	0.90100	1.00000	0.29701	0.78947	0.78311	1	0.65502	9
#15 multigs+multilearn TE	0.74475	0.79094	0.28294	0.18881	0.00205	0.83333	0.95023	0.95726	0.95079	1.00000	0.67011	9	0.76713	4
#16 multigs+multilearn+forgets TE	1.00000	0.96929	0.08701	0.10260	0.07512	0.80808	0.94966	0.95981	0.98770	1.00000	0.69393	2	0.80868	1
#17 multiprior TE	0.02349	0.01205	1.00000	0.78363	0.60840	1.00000	1.00000	0.99556	0.54218	0.84211	0.68074	6	0.56997	13
#18 multiprior+forgets TE	0.38713	0.41454	0.48855	0.81915	0.93068	0.00000	0.00000	0.00000	0.00000	0.00000	0.30400	17	0.21504	17

Table E15. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 15.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.10848	0.11232	0.86656	0.96022	1.00000	0.76141	0.87384	0.78735	0.20245	0.50000	0.61726	4	0.42520	14
#02 vanilla+forgets	0.27702	0.30704	0.76535	0.85244	0.47600	0.68513	0.75055	0.57958	0.18098	0.50000	0.53741	8	0.45259	13
#03 multigs T	0.38328	0.43642	0.15971	0.62885	0.10237	0.06275	0.10285	0.77286	0.94479	1.00000	0.45939	14	0.49782	8
#04 multigs+forgets T	0.45212	0.64089	0.00000	0.00000	0.00000	0.00000	0.00000	0.58667	1.00000	1.00000	0.36797	17	0.51550	7
#05 multilearn T	0.15994	0.17780	0.99266	0.90966	0.87469	0.83929	1.00000	1.00000	0.10123	0.50000	0.65553	1	0.46182	10
#06 multilearn+forgets T	0.26085	0.27773	0.85972	0.87031	0.91550	0.76141	0.87384	0.78735	0.10736	0.50000	0.62141	3	0.46118	11
#07 multigs+multilearn T	0.37899	0.43758	0.16071	0.62439	0.10203	0.06275	0.10285	0.77286	0.94479	1.00000	0.45870	15	0.49747	9
#08 multigs+multilearn+forgets T	0.51012	0.59427	0.13005	0.70939	0.05488	0.06275	0.10285	0.77286	0.95399	1.00000	0.48912	13	0.54186	5
#09 multiprior T	0.00000	0.00000	0.83004	0.74560	0.90441	0.76141	0.87384	0.78735	0.00000	0.00000	0.49027	12	0.26524	18
#10 multiprior+forgets T	0.21588	0.13747	0.63919	0.10514	0.56062	0.53714	0.51218	0.84984	0.07975	0.00000	0.36372	18	0.26824	17
#11 multigs TE	0.88811	0.93218	0.15387	0.55999	0.15252	0.12668	0.20783	0.33363	0.92025	1.00000	0.52751	9	0.67018	3
#12 multigs+forgets TE	0.94580	1.00000	0.15165	0.62492	0.09361	0.06275	0.10285	0.77286	0.94479	1.00000	0.56992	7	0.68416	2
#13 multilearn TE	0.15898	0.17850	0.98424	0.89850	0.89312	0.83929	1.00000	1.00000	0.10123	0.50000	0.65539	2	0.46037	12
#14 multilearn+forgets TE	0.28149	0.27553	1.00000	1.00000	0.74774	1.00000	0.86000	0.00000	0.09509	0.50000	0.57599	5	0.52535	6
#15 multigs+multilearn TE	0.87527	0.91757	0.15156	0.56833	0.14094	0.12668	0.20783	0.33363	0.92331	1.00000	0.52451	10	0.66573	4
#16 multigs+multilearn+forgets TE	1.00000	0.91202	0.19298	0.69026	0.06632	0.06275	0.10285	0.77286	0.94785	1.00000	0.57479	6	0.68593	1
#17 multiprior TE	0.06900	0.05575	0.83004	0.74560	0.90441	0.76141	0.87384	0.78735	0.00000	0.00000	0.50274	11	0.28603	16
#18 multiprior+forgets TE	0.29592	0.19230	0.63919	0.10514	0.56062	0.53714	0.51218	0.84984	0.07975	0.00000	0.37721	16	0.29072	15

Table E16. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 16.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.00000	0.23640	0.71396	0.99935	0.48975	0.91667	0.82126	0.69697	0.58185	0.94444	0.64006	3	0.56555	3
#02 vanilla+forgets	0.61397	0.55942	1.00000	0.72207	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.38955	10	0.36223	14
#03 multigs T	0.10367	0.07970	0.02807	0.92368	0.00469	0.62500	1.00000	0.96858	0.96797	1.00000	0.57014	7	0.46740	9
#04 multigs+forgets T	0.83749	0.75171	0.74265	0.23693	0.56644	0.00000	0.00000	0.00000	0.00000	0.00000	0.31352	16	0.38864	12
#05 multilearn T	0.00568	0.21768	0.71502	1.00000	0.49121	0.91667	0.82126	0.69697	0.58185	0.94444	0.63908	4	0.56356	4
#06 multilearn+forgets T	0.66199	0.53808	0.94771	0.59640	0.92307	0.00000	0.00000	0.00000	0.00000	0.00000	0.36672	11	0.35796	15
#07 multigs+multilearn T	0.10995	0.08471	0.02984	0.92565	0.00000	0.62500	1.00000	0.96858	0.96797	1.00000	0.57117	6	0.46958	8
#08 multigs+multilearn+forgets T	0.74924	0.68761	0.75353	0.17865	0.58840	0.00000	0.00000	0.00000	0.00000	0.00000	0.29574	17	0.36506	13
#09 multiprior T	0.10000	0.00000	0.89207	0.65342	0.12820	1.00000	0.66371	0.57607	0.37367	0.83333	0.52205	9	0.53318	7
#10 multiprior+forgets T	0.57220	0.18404	0.92107	0.72458	0.94752	0.00000	0.00000	0.00000	0.00000	0.00000	0.33482	14	0.27934	18
#11 multigs TE	0.39350	0.63430	0.00077	0.75959	0.23325	0.61111	0.98551	1.00000	1.00000	1.00000	0.66180	2	0.60661	1
#12 multigs+forgets TE	0.90212	0.92535	0.62464	0.00000	0.40821	0.00000	0.00000	0.00000	0.00000	0.00000	0.28602	18	0.40866	11
#13 multilearn TE	0.00499	0.17261	0.71484	0.99992	0.49103	0.91667	0.82126	0.69697	0.58185	0.94444	0.63446	5	0.55590	6
#14 multilearn+forgets TE	0.63984	0.58361	0.91212	0.60772	0.90458	0.00000	0.00000	0.00000	0.00000	0.00000	0.36479	12	0.35593	16
#15 multigs+multilearn TE	0.39460	0.63330	0.00000	0.76252	0.23503	0.61111	0.98551	1.00000	1.00000	1.00000	0.66221	1	0.60650	2
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.66281	0.09883	0.42953	0.00000	0.00000	0.00000	0.00000	0.00000	0.31912	15	0.44380	10
#17 multiprior TE	0.19275	0.08023	0.89207	0.65342	0.12820	1.00000	0.66371	0.57607	0.37367	0.83333	0.53934	8	0.56201	5
#18 multiprior+forgets TE	0.66832	0.22597	0.92107	0.72458	0.94752	0.00000	0.00000	0.00000	0.00000	0.00000	0.34887	13	0.30277	17

Table E17. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 17.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.22816	0.00446	0.95696	0.95118	0.63982	0.99218	1.00000	1.00000	0.34777	0.60000	0.67205	2	0.52159	9
#02 vanilla+forgets	0.45740	0.29415	0.90811	0.89352	0.21017	1.00000	0.79560	0.67984	0.43317	0.80000	0.64720	6	0.64881	5
#03 multigs T	0.20619	0.16816	0.54723	0.74521	0.98216	0.30973	0.33497	0.90702	0.83787	1.00000	0.60386	9	0.51153	13
#04 multigs+forgets T	0.35382	0.51753	0.39520	0.61013	0.88548	0.21739	0.17127	0.63529	0.87995	1.00000	0.56661	13	0.56065	8
#05 multilearn T	0.22157	0.00698	0.95683	0.95161	0.64005	0.99218	1.00000	1.00000	0.34777	0.60000	0.67170	4	0.52089	11
#06 multilearn+forgets T	0.42960	0.31562	0.94236	0.86267	0.51905	0.85106	0.67368	0.85466	0.29332	0.60000	0.63420	7	0.57199	7
#07 multigs+multilearn T	0.20788	0.16716	0.54895	0.72186	1.00000	0.30973	0.33497	0.90702	0.83787	1.00000	0.60355	10	0.51193	12
#08 multigs+multilearn+forgets T	0.44874	0.22221	0.00000	0.00000	0.29600	0.00000	0.00000	0.00000	1.00000	1.00000	0.29670	18	0.44516	14
#09 multiprior T	0.00000	0.00000	0.99876	0.97398	0.13416	0.82794	0.54030	0.54217	0.00000	0.00000	0.40173	17	0.30445	18
#10 multiprior+forgets T	0.17731	0.12757	1.00000	1.00000	0.16303	0.90254	0.43512	0.37391	0.02599	0.00000	0.42055	15	0.37223	16
#11 multigs TE	0.72554	0.67116	0.69397	0.71691	0.00000	0.50459	0.44608	0.23514	0.71658	1.00000	0.57100	12	0.71864	4
#12 multigs+forgets TE	0.86502	0.96549	0.61961	0.56875	0.34624	0.40541	0.27395	0.37391	0.78465	1.00000	0.62030	8	0.77336	2
#13 multilearn TE	0.22508	0.00658	0.95698	0.95178	0.64011	0.99218	1.00000	1.00000	0.34777	0.60000	0.67205	3	0.52143	10
#14 multilearn+forgets TE	0.45744	0.29998	0.94530	0.88379	0.51876	0.85106	0.67368	0.85466	0.34901	0.80000	0.66337	5	0.61713	6
#15 multigs+multilearn TE	0.72651	0.67069	0.69490	0.71392	0.00534	0.50459	0.44608	0.23514	0.71658	1.00000	0.57138	11	0.71888	3
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.76793	0.92362	0.83842	0.60748	0.62509	0.51561	0.65347	1.00000	0.79316	1	0.83815	1
#17 multiprior TE	0.03944	0.04974	0.99876	0.97398	0.13416	0.82794	0.54030	0.54217	0.00000	0.00000	0.41065	16	0.31931	17
#18 multiprior+forgets TE	0.23113	0.19164	1.00000	1.00000	0.16303	0.90254	0.43512	0.37391	0.02599	0.00000	0.43234	14	0.39188	15

Table E18. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 19.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.09126	0.09011	0.83066	0.72922	0.21076	0.90000	0.93399	0.93399	0.76800	0.94737	0.64354	4	0.60457	4
#02 vanilla+forgets	0.57678	0.65115	0.21918	0.08983	0.70239	0.00000	0.00000	0.00000	0.00000	0.00000	0.22393	16	0.24118	16
#03 multigs T	0.12701	0.03476	0.10720	0.79452	0.41173	0.76271	1.00000	1.00000	1.00000	1.00000	0.62379	8	0.50528	10
#04 multigs+forgets T	0.63460	0.72377	0.31007	0.11286	0.76881	0.00000	0.00000	0.00000	0.00000	0.00000	0.25501	14	0.27807	13
#05 multilearn T	0.09337	0.09040	0.82624	0.73077	0.20978	0.90000	0.93399	0.93399	0.76800	0.94737	0.64339	6	0.60423	6
#06 multilearn+forgets T	0.57280	0.64963	0.42507	0.17965	0.78850	0.00000	0.00000	0.00000	0.00000	0.00000	0.26157	13	0.27458	14
#07 multigs+multilearn T	0.12638	0.03487	0.11536	0.79676	0.41001	0.76271	1.00000	1.00000	1.00000	1.00000	0.62461	7	0.50656	9
#08 multigs+multilearn+forgets T	0.54617	0.71344	0.69309	0.33751	0.97990	0.00000	0.00000	0.00000	0.00000	0.00000	0.32701	12	0.32545	12
#09 multiprior T	0.00000	0.00000	1.00000	0.55253	0.00000	1.00000	0.88011	0.83820	0.54400	0.84211	0.56569	10	0.56435	8
#10 multiprior+forgets T	0.53280	0.46614	0.00000	0.00000	0.61189	0.00000	0.00000	0.00000	0.00000	0.00000	0.16108	18	0.16649	18
#11 multigs TE	0.50666	0.54109	0.10877	0.79578	0.41275	0.76271	1.00000	1.00000	1.00000	1.00000	0.71278	2	0.65320	2
#12 multigs+forgets TE	0.96089	0.98775	0.74547	0.37085	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.40650	11	0.44902	11
#13 multilearn TE	0.09335	0.09063	0.82647	0.73069	0.20980	0.90000	0.93399	0.93399	0.76800	0.94737	0.64343	5	0.60430	5
#14 multilearn+forgets TE	0.58535	0.68399	0.25336	0.12245	0.72678	0.00000	0.00000	0.00000	0.00000	0.00000	0.23719	15	0.25378	15
#15 multigs+multilearn TE	0.50523	0.53934	0.10606	0.79401	0.41176	0.76271	1.00000	1.00000	1.00000	1.00000	0.71191	3	0.65222	3
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.42129	1.00000	0.54288	0.90000	0.97291	0.97291	0.79733	1.00000	0.86073	1	0.85310	1
#17 multiprior TE	0.03928	0.06668	1.00000	0.55253	0.00000	1.00000	0.88011	0.83820	0.54400	0.84211	0.57629	9	0.58201	7
#18 multiprior+forgets TE	0.60404	0.48967	0.00000	0.00000	0.61189	0.00000	0.00000	0.00000	0.00000	0.00000	0.17056	17	0.18228	17

Table E19. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 21.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.19197	0.29397	0.64411	0.56494	0.75961	0.47475	1.00000	1.00000	0.74725	1.00000	0.66766	2	0.55867	9
#02 vanilla+forgets	0.39726	0.60604	0.85528	0.26299	0.49762	0.57551	0.30823	0.72431	0.52473	0.66667	0.54186	9	0.60425	7
#03 multigs T	0.51056	0.42034	0.16918	0.13366	0.65317	0.09126	0.50254	0.43229	0.94231	1.00000	0.48553	13	0.52227	13
#04 multigs+forgets T	0.66673	0.66087	0.00000	0.07329	0.87483	0.00000	0.12723	0.69591	1.00000	1.00000	0.50988	12	0.55460	11
#05 multilearn T	0.16730	0.30564	0.63811	0.53125	0.76121	0.47475	1.00000	1.00000	0.74725	1.00000	0.66255	3	0.55551	10
#06 multilearn+forgets T	0.36610	0.54484	0.82009	0.00000	0.45928	0.57551	0.30823	0.72431	0.65659	1.00000	0.54550	8	0.66052	2
#07 multigs+multilearn T	0.51057	0.41894	0.16918	0.13358	0.65313	0.09126	0.50254	0.43229	0.94231	1.00000	0.48538	14	0.52204	14
#08 multigs+multilearn+forgets T	0.83071	0.55425	0.03815	0.26824	0.95389	0.00000	0.12723	0.69591	1.00000	1.00000	0.54684	7	0.57052	8
#09 multiprior T	0.00000	0.00000	0.99788	0.25515	0.08393	0.89053	0.40798	0.29825	0.06319	0.00000	0.29969	18	0.32527	18
#10 multiprior+forgets T	0.19089	0.25584	1.00000	1.00000	0.00000	1.00000	0.82534	0.00000	0.00000	0.00000	0.42721	16	0.40779	16
#11 multigs TE	0.78447	0.70620	0.16937	0.13338	0.65302	0.09126	0.50254	0.43229	0.94231	1.00000	0.54148	10	0.61560	5
#12 multigs+forgets TE	0.92421	1.00000	0.00041	0.07453	0.87479	0.00000	0.12723	0.69591	1.00000	1.00000	0.56971	5	0.65410	3
#13 multilearn TE	0.16407	0.25944	0.63668	0.52383	0.76150	0.47475	1.00000	1.00000	0.74725	1.00000	0.65675	4	0.54703	12
#14 multilearn+forgets TE	0.36124	0.61320	0.58875	0.37595	0.83651	0.56667	0.00000	0.58922	0.57967	1.00000	0.55112	6	0.61825	4
#15 multigs+multilearn TE	0.78447	0.70519	0.16936	0.13339	0.65299	0.09126	0.50254	0.43229	0.94231	1.00000	0.54138	11	0.61543	6
#16 multigs+multilearn+forgets TE	1.00000	0.96462	0.39074	0.14712	1.00000	0.16800	0.60320	0.85312	0.76374	1.00000	0.68905	1	0.71452	1
#17 multiprior TE	0.02382	0.06271	0.99788	0.25515	0.08393	0.89053	0.40798	0.29825	0.06319	0.00000	0.30834	17	0.33969	17
#18 multiprior+forgets TE	0.21654	0.32925	1.00000	1.00000	0.00000	1.00000	0.82534	0.00000	0.00000	0.00000	0.43711	15	0.42430	15

Table E20. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 23.

BKT model	RMSE	AUC	Correlation			F1			Average	Ideal path	Composite score A	Rank A	Composite score B	Rank B
			Formative	Midterm	Final	Formative	Midterm	Final						
#01 vanilla	0.12576	0.15467	0.48796	0.73770	0.87946	0.38500	0.00000	0.38356	0.59438	0.66667	0.44152	13	0.40241	16
#02 vanilla+forgets	0.49501	0.54745	0.53706	0.90185	0.90678	0.58481	1.00000	0.58333	0.59036	0.66667	0.68133	2	0.57023	7
#03 multigs T	0.36295	0.55657	0.02284	0.17861	0.07298	0.00000	0.88757	0.00000	1.00000	1.00000	0.40815	18	0.49039	11
#04 multigs+forgets T	0.53576	0.73683	0.02843	0.15845	0.09902	0.00000	0.88757	0.00000	1.00000	1.00000	0.44461	12	0.55017	9
#05 multilearn T	0.12645	0.16659	0.49791	0.78318	0.84711	0.38500	0.00000	0.38356	0.59438	0.66667	0.44508	11	0.40617	15
#06 multilearn+forgets T	0.49991	0.51613	0.55094	1.00000	0.80741	0.58481	1.00000	0.58333	0.57430	0.66667	0.67835	3	0.56546	8
#07 multigs+multilearn T	0.36481	0.55804	0.04686	0.26168	0.00000	0.00000	0.88757	0.00000	1.00000	1.00000	0.41190	17	0.49495	10
#08 multigs+multilearn+forgets T	0.62284	0.74678	0.07265	0.29255	0.01689	0.00000	0.88757	0.00000	1.00000	1.00000	0.46393	10	0.57371	6
#09 multiprior T	0.00000	0.00000	0.88699	0.00000	0.87034	1.00000	0.07903	1.00000	0.18876	0.33333	0.43585	16	0.40151	17
#10 multiprior+forgets T	0.32327	0.33883	1.00000	0.51424	1.00000	1.00000	0.07903	1.00000	0.00000	0.00000	0.52554	6	0.44368	13
#11 multigs TE	0.79675	0.85036	0.00000	0.11487	0.11882	0.00000	0.88757	0.00000	1.00000	1.00000	0.47684	9	0.60785	4
#12 multigs+forgets TE	0.93096	0.96702	0.02818	0.15680	0.09913	0.00000	0.88757	0.00000	1.00000	1.00000	0.50697	7	0.65436	2
#13 multilearn TE	0.12540	0.13744	0.48826	0.73890	0.87873	0.38500	0.00000	0.38356	0.59438	0.66667	0.43983	15	0.39953	18
#14 multilearn+forgets TE	0.50635	0.58355	0.55151	0.97465	0.89262	0.58481	1.00000	0.58333	0.58233	0.66667	0.69261	1	0.57925	5
#15 multigs+multilearn TE	0.79556	0.85310	0.01242	0.14381	0.09666	0.00000	0.88757	0.00000	1.00000	1.00000	0.47891	8	0.61018	3
#16 multigs+multilearn+forgets TE	1.00000	1.00000	0.05383	0.19484	0.15242	0.00000	0.88757	0.00000	0.99598	1.00000	0.52846	5	0.67497	1
#17 multiprior TE	0.02235	0.02120	0.88699	0.00000	0.87034	1.00000	0.07903	1.00000	0.18876	0.33333	0.44020	14	0.40877	14
#18 multiprior+forgets TE	0.34864	0.35197	1.00000	0.51424	1.00000	1.00000	0.07903	1.00000	0.00000	0.00000	0.52939	4	0.45010	12

Table E21. BKT model ranking (normalised features, KM estimation based on the testing subset) – DKC 25.

APPENDIX F

Learning experience questionnaire used for Self-Practice and Controlled Environment formative assessments

Please evaluate your learning experience using a scale from 1 to 7:

1. I do not feel comfortable during the formative assessments.	1	2	3	4	5	6	7	I feel comfortable during the formative assessments.
2. I feel that the formative assessments had a negative impact on my learning performance.	1	2	3	4	5	6	7	I feel that the formative assessments had a positive impact on my learning performance.
3. I feel that the formative assessments have not helped me to improve my knowledge of basic programming concepts.	1	2	3	4	5	6	7	I feel that the formative assessments have helped me to improve my knowledge of basic programming concepts.
4. I would not recommend using formative assessments in the course.	1	2	3	4	5	6	7	I would recommend using formative assessments in the course.
5. I would not recommend using adaptive formative assessments that last shorter.	1	2	3	4	5	6	7	I would recommend using adaptive formative assessments that last shorter.
6. I would not recommend using bonus points in the course based on formative assessments.	1	2	3	4	5	6	7	I would recommend using bonus points in the course based on formative assessments.
7. I would not recommend using self-practice formative assessments one day before laboratory exercises.	1	2	3	4	5	6	7	I would recommend using self-practice formative assessments one day before laboratory exercises.

APPENDIX G

System Usability Scale (SUS) questionnaire

Please evaluate the formative assessment programming environment using a scale from 1 to 5:

1. I think that I would like to use the formative assessment programming environment frequently.	1	2	3	4	5
2. I found the formative assessment programming environment unnecessarily complex.	1	2	3	4	5
3. I thought the formative assessment programming environment was easy to use.	1	2	3	4	5
4. I think I would need a technical person's support to use this formative assessment programming environment.	1	2	3	4	5
5. I found the various functions in this formative assessment programming environment were well integrated.	1	2	3	4	5
6. I thought there was too much inconsistency in this formative assessment programming environment.	1	2	3	4	5
7. I would imagine that most people would learn to use this formative assessment programming environment very quickly.	1	2	3	4	5
8. I found the formative assessment programming environment very cumbersome to use.	1	2	3	4	5
9. I felt very confident using the formative assessment programming environment.	1	2	3	4	5
10. I needed to learn many things before I could get going with this formative assessment programming environment.	1	2	3	4	5

Curriculum Vitae

Ines Šarić-Grgić

Ines Šarić-Grgić was born in 1985 in Split, Croatia. She completed her secondary education at the III Gymnasium in Split before obtaining a Master's degree in Computing from the Faculty of Electrical Engineering, Mechanical Engineering, and Naval Architecture at the University of Split in 2008. In 2013, she obtained a specialist degree in Business Economics from the Faculty of Economics at the same university.

Since 2014, she has been a teaching and research assistant at the Faculty of Science, University of Split. Her research focuses on educational data mining, learning analytics and intelligent tutoring systems. She has contributed to two scientific projects on Adaptive Courseware based on Natural Language Processing (AC&NL Tutor, Office of Naval Research, USA). She has presented her research at several international conferences, including Intelligent Tutoring Systems (2016, 2018, 2019), Computer Supported Education (2017), Software, Telecommunications and Computer Networks (2018, 2019) and Adaptive Instructional Systems (2020).

She is married to Lovre and is the mother of Zoja.

Životopis

Ines Šarić-Grgić

Ines Šarić-Grgić rođena je 1985. godine u Splitu. Srednjoškolsko obrazovanje završila je u III. gimnaziji u Splitu, nakon čega je 2008. godine stekla diplomu magistre računarstva na Fakultetu elektrotehnike, strojarstva i brodogradnje Sveučilišta u Splitu. Godine 2013. završila je specijalistički studij Poslovne ekonomije na Ekonomskom fakultetu istog sveučilišta.

Od 2014. godine zaposlena je kao asistentica i stručna suradnica na Prirodoslovno-matematičkom fakultetu Sveučilišta u Splitu. Njezino istraživačko područje obuhvaća rudarenje podataka u obrazovanju, analitiku učenja i inteligentne tutorske sustave. Sudjelovala je u dva znanstvena projekta (Adaptive Courseware based on Natural Language Processing - AC&NL Tutor, Office of Naval Research, SAD). Rezultate istraživanja predstavila je na nekoliko međunarodnih konferencija, uključujući Intelligent Tutoring Systems (2016, 2018, 2019), Computer-Supported Education (2017), Software, Telecommunications and Computer Networks (2018, 2019) i Adaptive Instructional Systems (2020).

U braku je s Lovrom i majka je Zoje.