

# Equity and Fairness of Bayesian Knowledge Tracing

Sebastian Tschatschek  
University of Vienna  
Faculty of Computer Science  
Research Network Data  
Science  
Vienna, Austria

Maria Knobelsdorf  
University of Vienna  
Faculty of Computer Science  
Computer Science Education  
Vienna, Austria

Adish Singla  
Max-Planck Institute for  
Software Systems  
Saarbrücken, Germany

## ABSTRACT

We consider the equity and fairness of curricula derived from Knowledge Tracing models. We begin by defining a unifying notion of an equitable tutoring system as a system that achieves maximum possible knowledge in minimal time for each student interacting with it. Realizing perfect equity requires tutoring systems that can provide individualized curricula per student. In particular, we investigate the design of equitable tutoring systems that derive their curricula from Knowledge Tracing models. We first show that the classical Bayesian Knowledge Tracing (BKT) model and their derived curricula can fall short of achieving equitable tutoring. To overcome this issue, we then propose a novel model, Bayesian-Bayesian Knowledge Tracing (B<sup>2</sup>KT), that naturally allows online individualization. We demonstrate that curricula derived from our model are more effective and equitable than those derived from existing models. Furthermore, we highlight that improving models with a focus on the fairness of next-step predictions can be insufficient to develop equitable tutoring systems.

## Keywords

equity & fairness, knowledge tracing, intelligent tutoring

## 1. INTRODUCTION

In recent years Massive Open Online Courses (MOOCs) and online educational platforms have gained significant importance. They hold the opportunity of providing education at scale and making education accessible to a larger part of the world's population. To facilitate learning in online education and enable customized learning paths for all students, intelligent tutoring systems can be employed while limiting the amount of manual work necessary for each student [11].

In that context, moving education from an offline setting to an online setting, has the potential to promote *Inclusion, Diversity, Equity, and Accessibility (IDEA)*. In particular, by reducing personnel efforts for tutoring, there is the opportunity to include students with diverse backgrounds and

skills, and, importantly, to support their learning equitably. To achieve this, an intelligent tutoring system must be able to adapt to the specific characteristics of each student.

While individualized tutoring has been studied in the community for many years, we consider individualization with a focus on equitable and fair tutoring in this paper. We start by providing a unifying definition of an equitable tutoring system. Our definition is based on the ethical principles of *beneficence* (“do the best”) and *non-maleficence* (“do not harm”) which are commonly adopted in bioethics and medical applications [1]. These principles dictates that we should provide tutoring which maximizes the achieved knowledge while minimizing a student's efforts. In particular we focus on modifying Bayesian Knowledge Tracing (BKT) [2] to better realize these ethical principles. To this end, we propose the Bayesian-Bayesian Knowledge Tracing (B<sup>2</sup>KT) model and demonstrate its advantages for equitable tutoring in several experiments. Furthermore, we investigate the relation of the commonly considered AUC score concerning the derived tutoring policies, finding that even if a BKT model appears fair in terms of the AUC score, the derived tutoring policies can be inequitable.

In summary, we make the following contributions: *(i)* We propose a unifying definition of equitable tutoring motivated by ethical principles. *(ii)* We propose the B<sup>2</sup>KT model which allows for effective individualization and demonstrate its benefits concerning equitable tutoring. *(iii)* We highlight that focusing on equity in terms of AUC can be insufficient to ensure equitable tutoring in terms of our definition.

An longer version of this paper with additional experimental results and extended discussion is available [15].

## 2. RELATED WORK

**Fairness in online education and BKT.** Several works have considered fairness in data-driven educational systems and intelligent tutoring, e.g., [7, 4, 17, 8]. In [7], the authors discussed implications of using data-driven predictive models for supporting education on fairness. They identified sources of bias and discrimination in “the process of developing and deploying these systems”, and discussed high-level possibilities to improve fairness of systems in the “action step”. In [8, 17], it was investigated how different data sources can provide helpful information to predict students' success in education. Key insights were that different data sources can help to make better predictions but have different characteristics in whether they over- or underestimate students' success [17], and that such predictions can include gender and

S. Tschatschek, M. Knobelsdorf, and A. Singla. Equity and fairness of Bayesian knowledge tracing. In A. Mitrovic and N. Bosch, editors, *Proceedings of the 15th International Conference on Educational Data Mining*, pages 578–582, Durham, United Kingdom, July 2022. International Educational Data Mining Society.

racial bias in some fairness measures which can be partly alleviated through post-hoc adjustments [8]. In [4] fairness in the context of BKT was studied, and it was found that tutoring policies basing on inaccurate BKT models can be inequitable, when considering the difference in learning success for different subpopulations as a measure of unfairness. Related work also considers adopting a Bayesian perspective for realizing fair decision rules under model uncertainty [3] and fairness in the context of non-i.i.d. data [19].

**Individualization in BKT.** Several papers have studied individualization of BKT models per student, e.g., [9, 10, 18]. In [10] the *prior per student* model was introduced which uses a student-specific parameter characterizing the students’ individual knowledge. [18] considered individualization through defining student and skill specific parameters which are fitted through gradient descent.

**Instructional policies.** Key for achieving equity according to our definition are instructional policies which stop practicing a skill at the right time. This problem has for instance been considered in [6, 12]. Further related work has investigated approaches leveraging deep models for creating policies to quickly assess students’ knowledge [16] and using reinforcement learning for optimizing tutoring policies [14, 5].

### 3. BACKGROUND & NOTATION

**Bayesian Knowledge Tracing.** Bayesian knowledge tracing (BKT) [2] is a model characterizing the skill acquisition process of students. For a single skill, it can be understood as a standard hidden Markov model in which the binary (latent) state encodes the mastery of the skill, and the binary observations indicate whether a practicing opportunity of the skill was solved correctly. Upon practicing a not yet mastered skill, the student acquires the skill with probability  $p(T)$ . Once a skill is mastered, it remains mastered. If a student has mastered the skill practiced by an exercise, they solve this exercise correctly with probability  $1 - p(S)$ . If a student has not mastered the skill, it guesses the correct answer with probability  $p(G)$ . At the beginning, a student has already mastered the skill with probability  $p(L_0)$ .

**Notation.** We consider the interaction of students  $s \in \mathcal{S}$  with an intelligent tutoring system. The interaction history up to time  $t$  is denoted as  $\mathcal{D}_t^s = \{(z_1, c_1), (z_2, c_2), \dots, (z_t, c_t)\}$ , where  $z_{t'} \in \mathcal{Z}$  is the skill practiced through an exercise at time  $t'$ ,  $c_{t'} \in \{0, 1\}$  is an indicator of whether the exercise was solved correctly, and  $\mathcal{Z}$  is the set of skills. In the context of BKT, we refer to the random variables (RVs) indicating whether skill  $i \in \mathcal{Z}$  is mastered at time  $t$  as  $Z_t^i$  and to the RVs indicating whether an exercise practicing that skill would be solved correctly as  $C_t^i$ . Sometimes we add another superscript  $s$  to indicate the student the RVs correspond to. Upper-case terms like  $Z_t^i$  denote RVs and their lower-case counterparts like  $z_t^i$  denote particular instantiations.

### 4. EQUITABLE TUTORING

In this section, we provide a definition of equity in intelligent tutoring and discuss its operationalization.

#### 4.1 Definition

We consider a tutoring setting in which a total of  $K$  skills ought to be taught to a set of students  $\mathcal{S}$  by an intelligent tutoring system employing a tutoring policy  $\pi: \mathcal{H} \rightarrow \mathcal{I} \cup \{\top\}$ . This policy maps histories  $h \in \mathcal{H}$  consisting of observations

of a student’s learning process to an exercise  $e \in \mathcal{I}$  to be practiced next or to a stop-action  $\top$ , which ends the teaching process. Each student can have different learning characteristics. Every tutoring policy  $\pi$  has an expected stopping time  $T^s(\pi)$ , i.e., the expected time of executing the stop action, and an expected knowledge  $L^s(\pi)$  acquired by the end of the teaching process, i.e.,  $L^s(\pi)$  is the expected number of mastered skills upon executing the stop action.

Our notion of equity is based on the ethical principles of *beneficence* and *non-maleficence*. We understand them to translate into the objective of maximizing a student’s knowledge using as little of the student’s resources as possible, i.e., performing a minimal number of exercises:

**Definition 1.** Consider a tutoring system employing a tutoring policy  $\pi$ . The policy  $\pi$  is equitable for student  $s$  iff

$$T^s(\pi) = \min_{\pi', L^s(\pi')=K} T^s(\pi') \quad \text{and} \quad L^s(\pi) = K.$$

A tutoring system is equitable if its tutoring policy is equitable for all students  $s \in \mathcal{S}$ .

Thus, informally, a tutoring system is equitable if it can teach all  $K$  skills in the minimal amount of time possible to any student. Note that our notion of equity is strongly related to that introduced in [4] (cf. discussion below). In the above definition, we implicitly assume that all students can master all  $K$  skills.<sup>1</sup> Importantly, a tutoring system can only be equitable if it is adaptive to the students which are interacting with it. In particular, it has to individualize the assignment of exercises and needs to carefully select the “stop action”, in order to achieve equity. The above definition describes an idealized notion of equity which in general cannot be achieved as the tutoring policy would have to teach using the optimal policy right from the beginning. Nevertheless, we can compare tutoring policies  $\pi$  in the spirit of the above definition. In particular, given two tutoring policies  $\pi$  and  $\pi'$  which both teach the same number of skills, we consider the policy  $\pi$  to be *more equitable* as compared to  $\pi'$  if for all students  $s \in \mathcal{S}$  it holds that  $T^s(\pi) \leq T^s(\pi')$ .

We note that our notion of equity is strongly related to that introduced in [4]. In [4], the authors “assume that an equitable outcome is when students from different demographics reach the same level of knowledge after receiving instruction”. The desideratum of achieving knowledge fast is later also added to their notion of equity whereas in our case it is a fundamental constituent. Furthermore, our interest extends to downstream implications of such a definition of equity, namely the individualization of knowledge tracing.

**Theoretical Implications.** Our definition of equity leads to the following (probably obvious) but important observation:

**Observation 1.** A tutoring system for a population of students with different learning characteristics can only be equitable if its tutoring policy is adaptive to the students.

Thus, we note that if the tutoring policy is derived deterministically from a non-adaptive, initially incorrect, model of the students, the tutoring system will in general not be equitable. Achieving equity would require basing a policy on rich side information in order to employ an optimal tutoring policy for each student right from the beginning. But such rich side information might not be available.

<sup>1</sup>Our definition can be easily generalized to account for an individual student’s maximal achievable knowledge.

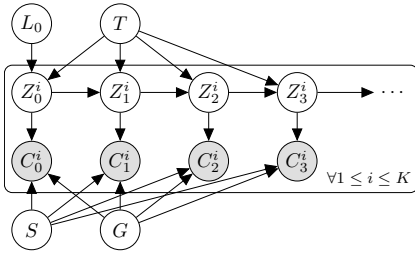


Figure 1: Graphical model of B<sup>2</sup>KT. The acquisition and application of the  $K$  skills depends on  $p(L_0)$ ,  $p(S)$ ,  $p(G)$ ,  $p(T)$ .

## 4.2 Operationalization

Tutoring policies are often either simple fixed strategies or derived from a model, e.g., a BKT model, such that each knowledge component is repeatedly exercised until it is mastered with a certain probability. But tutoring policies based on incorrect or non-adaptive models can result in a student not acquiring all skills or suggest to perform too many practicing opportunities. Thus the following two general directions are important for building equitable tutoring systems: (i) **Using side information.** Any available side information about a student should be used to individualize the underlying models. In the context of classical BKT models, the side information could be used to make an initial guess about the key parameters of the model ( $p(L_0), p(S), p(G), p(T)$ ). (ii) **Online adaptation.** Even when using side information, a model is likely not perfectly individualized to all students. To further adjust the models in such cases, online adaption of the models during interaction seems promising.

## 5. PROPOSED APPROACH: B<sup>2</sup>KT

In this section, we propose a Bayesian variant of the classical BKT model which enables online adaption to student’s parameters from which individualized — potentially more equitable — policies can be derived, cf. Figure 1.

We assume that each student  $s$  has its own learning dynamics, described by student-specific parameters  $\theta^s$ . If the learning dynamics can be described using a BKT model,  $\theta^s = (p(L_0^s), p(T^s), p(S^s), p(G^s))$ . We assume these learning dynamics to apply for the acquisition of *all* skills. In practice, we don’t know these parameters and need to infer them. To this end, we take a Bayesian approach, and we assume a set of possible parameters  $\Theta$  such that  $\theta^s \in \Theta$  and a prior distribution  $p_0(\theta^s)$ . Based on  $t$  observations of a student’s practicing exercises collected in  $\mathcal{D}_t$ , we can compute the probability that a student has mastered a specific skill and base tutoring policies thereon. As we don’t know  $\theta^s$ , this requires marginalizing out the (unknown) parameters  $\theta^s$ . In this way the different possible parameters and their influence for predicting the knowledge state get re-weighted according to the available data. In particular, we compute

$$p(Z_t^{s,i} | \mathcal{D}_t) = \int_{\theta \in \Theta} \underbrace{p(Z_t^{s,i} | \theta, \mathcal{D}_t)}_{=:(\#1)} \underbrace{p(\theta | \mathcal{D}_t)}_{=:(\#2)} d\theta, \quad (1)$$

where  $Z_t^{s,i}$  is a random variable indicating whether skill  $i$  is mastered at time  $t$  by student  $s$ . For only a few possible parameters  $\theta$ , the above equation can be solved exactly by enumeration and by observing that both terms (#1) and

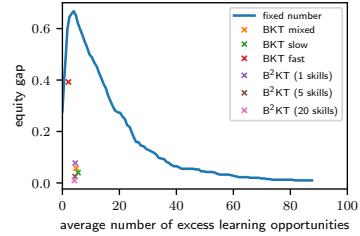


Figure 2: Equity gap vs number of excess learning opportunities. B<sup>2</sup>KT becomes more equitable as more skills are taught.

(#2) can be computed efficiently by the following recursion:

$$\begin{aligned} \alpha_0^\theta(l) &= p(Z_0^{s,i} = l | \theta) = p(L_0)^l (1 - p(L_0))^{1-l} \\ \alpha_{t+1}^\theta(l) &= p(Z_{t+1}^{s,i} = l, \mathbf{c}_{t+1}^i) \\ &= \sum_{z_t^{s,i}} p(\mathbf{c}_{t+1}^i | Z_{t+1}^{s,i} = l) p(Z_{t+1}^{s,i} = l | Z_t^{s,i} = z_t^{s,i}) \alpha_t^\theta(z_t^{s,i}) \end{aligned}$$

Here  $\mathbf{c}_t^i$  collects all observations with respect to practicing the  $i$ th skill up to time  $t$ , and  $c_t^i$  is the  $t$ ’th entry of  $\mathbf{c}_t^i$ . Then

$$\begin{aligned} (\#1) &= p(Z_t^{s,i} = 1 | \theta, \mathcal{D}_t) = \frac{\alpha_t^\theta(1)}{\alpha_t^\theta(0) + \alpha_t^\theta(1)}, \text{ and} \\ (\#2) &= p(\Theta = \theta | \mathcal{D}_t) = \frac{p_0(\theta) \cdot (\alpha_t^\theta(0) + \alpha_t^\theta(1))}{\sum_{\theta' \in \Theta} p_0(\theta') \cdot (\alpha_t^{\theta'}(0) + \alpha_t^{\theta'}(1))}. \end{aligned}$$

## 6. EXPERIMENTS

We perform experiments on synthetic data and consider settings in which the learning rate  $p(T^s)$  is assumed to be unknown. This is motivated by previous work which has identified the learning rate as a key parameter for improving BKT based models [18]. In all presented results we denote the average stopping time of a policy for a population of students by  $T_{\text{stop}}$  and the average number of acquired skills by % skills. We consider THRESHOLD( $\tau$ ) curricula based on knowledge tracing models. These curricula repeatedly exercise a skill until it is mastered with a probability of at least  $\tau$  under the model. We consider the following models: (i) BKT: the classical BKT models with fixed parameters; (ii) B<sup>2</sup>KT: the proposed Bayesian-BKT model.

**Students with different learning behaviors.** We study the equity of tutoring policies when the students are sampled uniformly from two groups, each containing students with learning dynamics described by a ground truth BKT model. In particular, we build on the experimental setup from [4] where there is a group of slow learners (BKT slow) and fast learners (BKT fast). In [4], the authors also fitted a BKT model to interaction data from students from both groups; we refer to the corresponding BKT model as BKT mixed. The parameters of the considered models are as follows:

	$p(L_0^s)$	$p(S^s)$	$p(G^s)$	$p(T^s)$
BKT slow	0.0	0.2	0.2	0.05
BKT fast	0.0	0.2	0.2	0.3
BKT mixed	0.071	0.203	0.209	0.096

We considered the interaction with 400 students, 200 from the slow and the fast group, respectively, and we compared the performance of THRESHOLD(0.95) tutoring policies based on these models for different numbers of skills that ought to be taught in Table 1. We observe that in the case of mismatch of the student properties and the BKT models used

**Table 1: Equity trade-offs of curricula derived from different models/parameterizations.**

THRESHOLD(0.95)	1 skill				5 skills				20 skills			
	slow learners		fast learners		slow learners		fast learners		slow learners		fast learners	
	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$
BKT slow	97.00	24.14	99.50	9.49	97.20	122.80	99.90	66.00	97.55	492.64	99.90	183.84
BKT fast	61.00	13.85	97.50	5.96	62.60	71.98	96.10	29.81	64.20	288.76	97.23	120.59
BKT mixed	95.00	23.51	100.00	8.33	95.40	113.67	99.90	40.93	94.53	466.55	99.68	169.86
B <sup>2</sup> KT	94.50	24.04	100.00	7.88	97.70	120.87	98.40	32.61	96.68	493.00	96.66	120.05

for the threshold policy, either only a small fraction of the skills (clearly below 95 %) is acquired or that more than necessary time is spent exercising. The mismatch issue is alleviated in the case of the B<sup>2</sup>KT model (assuming a uniform prior over both types of students), in particular for a larger number of skills. Intuitively this is because, in the case of multiple skills, the model has more opportunities to learn about the students’ characteristics and leverage this knowledge in later tutoring. This fact is also illustrated in Figure 2 in which we reproduce and extend an experiment from [4] in which we compare the “equity gap” (the difference in the percentage of skills mastered by fast and slow students, respectively) to the number of excess learning opportunities. Importantly, B<sup>2</sup>KT becomes more equitable as more skills are taught.

**Out-of-distribution generalization.** We test whether B<sup>2</sup>KT can help with aspects relevant to *inclusion and diversity*. In particular, we consider a stylized mismatch setting in which a tutoring system interacts with students who have a learning behavior not considered when building the system. In addition to the previous two types of students, we assume a third type of learner (BKT med) with the following parameters:  $p(L_0^s) = 0.0, p(S^s) = 0.2, p(G^s) = 0.2, p(T^s) = 0.18$ . We considered THRESHOLD(0.95) policies based on BKT models of slow and fast learners and the B<sup>2</sup>KT model with a uniform prior over slow and fast learners. Our results are presented in Table 2. We observe that the performance of the policies derived from the B<sup>2</sup>KT model have comparable performance to those derived from the true model (although the true model has zero posterior probability) whereas other models yield policies worse in terms of stopping at the right time or teaching the right amount of skills. This property of B<sup>2</sup>KT can be helpful for promoting inclusion, e.g., when interacting with students who were underrepresented in the data used for building an intelligent tutoring system.

**Fair next step predictions do not necessarily imply equitable tutoring.** We show empirically that models which might appear to be fair when looking at their AUCs for different groups of students do not necessarily yield equitable tutoring policies. In particular, we again focus on a student population consisting of two groups of students:

	$p(L_0^s)$	$p(S^s)$	$p(G^s)$	$p(T^s)$
Group 1	0.0	0.1	0.4	0.1
Group 2	0.0	0.1	0.2	0.3

We generated data of 400 students (50% from group 1 and group 2, respectively) in a setting with 20 skills and 1000 random exercises from a BKT model. The true model of group 1’s students achieved an AUC of 0.7393 for group 1’s students, while the true model of group 2’s students achieved an AUC of 0.6710 for group 2’s students.

Looking only at the AUC, the two models appear rather inequitable (there is no group parity). Thus it might appear

sensible to aim to use a BKT model for tutoring which has comparable AUCs for both groups in order to promote equity. For instance, a BKT model using parameters  $p(L_0) = 0, p(S) = 0.4, p(G) = 0.1, p(T) = 0.65$  achieves an AUC of 0.6719 on group 1’s students and of 0.6733 on group 2’s students, respectively. That is, the AUCs on the two groups are approximately equal. However, when looking at the different models with respect to their tutoring performance using a THRESHOLD(0.95)-policy, we observe a very different picture, cf. Table 3. In particular, the fraction of skills taught differs significantly between the two groups: In group 1 only 28.68% of the skills are acquired by the students on average while in group 2 74.70% of the skills are acquired. This finding is closely related to the observation that models with greatly different characteristics can have similar AUCs [13].

**Table 2: Out-of-distribution generalization.**

THRESHOLD(0.95)	1 skill		5 skills		20 skills	
	BKT med		BKT med		BKT med	
	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$	% skills	$T_{\text{stop}}$
BKT slow	99.50	11.28	99.55	56.45	99.59	225.25
BKT fast	90.75	7.61	91.55	37.59	91.70	151.36
BKT mixed	99.50	10.46	99.00	51.71	99.21	211.29
BKT med	98.25	8.82	97.35	45.59	97.84	184.20
B <sup>2</sup> KT	98.75	10.33	97.50	48.80	94.19	168.36

**Table 3: Fairness in terms of similar AUCs on different groups does not imply fairness in terms of the derived curricula.**

group	group fair wrt AUC			true model wrt group		
	AUC	% skills	$T_{\text{stop}}$	AUC	% skills	$T_{\text{stop}}$
group 1	0.6719	28.68	61	0.7393	96.13	308
group 2	0.6733	74.70	64	0.6710	96.35	105

## 7. CONCLUSION

We considered the equity and fairness of curricula derived from knowledge tracing models, and provided a unifying definition of equitable tutoring systems. Our definition is, in many practical settings, not realizable but suggests that the individualization of tutoring policies to students is key for realizing equity. We proposed the B<sup>2</sup>KT model, a Bayesian variant of the classical BKT model, and demonstrated in various experiments that it can be beneficial for realizing equitable tutoring systems and promoting IDEA more generally. Furthermore, we highlighted that improving and evaluating models with the main focus on next-step predictions can be insufficient to develop equitable tutoring systems.

## 8. ACKNOWLEDGMENTS

Adish Singla acknowledges support by the European Research Council (ERC) under the Horizon Europe programme (ERC StG, grant agreement No. 101039090).

Sebastian Tschiatschek acknowledges funding by the Vienna Science and Technology Fund (WWTF) and the City of Vienna through project ICT20-058.

## 9. REFERENCES

- [1] T. L. Beauchamp, J. F. Childress, et al. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.
- [2] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.
- [3] C. Dimitrakakis, Y. Liu, D. C. Parkes, and G. Radanovic. Bayesian fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):509–516, 2019.
- [4] S. Doroudi and E. Brunskill. Fairer but not fair enough on the equitability of knowledge tracing. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 335–339, 2019.
- [5] J. He-Yueya and A. Singla. Quizzing policy using reinforcement learning for inferring the student knowledge state. *International Educational Data Mining Society*, 2021.
- [6] T. Käser, S. Klingler, and M. Gross. When to stop? towards universal instructional policies. In *Proceedings of the sixth international conference on learning analytics & knowledge*, pages 289–298, 2016.
- [7] R. F. Kizilcec and H. Lee. Algorithmic fairness in education. *arXiv preprint arXiv:2007.05443*, 2020.
- [8] H. Lee and R. F. Kizilcec. Evaluation of fairness trade-offs in predicting student success. *arXiv preprint arXiv:2007.00088*, 2020.
- [9] J. I. Lee and E. Brunskill. The impact on individualizing student models on necessary practice opportunities. *International Educational Data Mining Society*, 2012.
- [10] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International conference on user modeling, adaptation, and personalization*, pages 255–266. Springer, 2010.
- [11] G. Paviotti, P. G. Rossi, and D. Zarka. Intelligent tutoring systems: an overview. *Pensa Multimedia*, pages 1–176, 2012.
- [12] R. Pelánek. Conceptual issues in mastery criteria: Differentiating uncertainty and degrees of knowledge. In *International Conference on Artificial Intelligence in Education*, pages 450–461. Springer, 2018.
- [13] R. Pelánek. The details matter: methodological nuances in the evaluation of student models. *User Modeling and User-Adapted Interaction*, 28(3):207–235, 2018.
- [14] A. Singla, A. N. Rafferty, G. Radanovic, and N. T. Heffernan. Reinforcement learning for education: Opportunities and challenges. *arXiv preprint arXiv:2107.08828*, 2021.
- [15] S. Tschitschek, M. Knobelsdorf, and A. Singla. Equity and Fairness of Bayesian Knowledge Tracing. *arXiv preprint arXiv:2205.02333*, 2022.
- [16] Z. Wang, S. Tschitschek, S. Woodhead, J. M. Hernández-Lobato, S. Peyton Jones, R. G. Baraniuk, and C. Zhang. Educational question mining at scale: Prediction, analysis and personalization. *AAAI Conference on Artificial Intelligence*, 35(17):15669–15677, 2021.
- [17] R. Yu, Q. Li, C. Fischer, S. Doroudi, and D. Xu. Towards accurate and fair prediction of college success: Evaluating different sources of student data. *International Educational Data Mining Society*, 2020.
- [18] M. V. Yudelson, K. R. Koedinger, and G. J. Gordon. Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education*, pages 171–180. Springer, 2013.
- [19] W. Zhang, J. C. Weiss, S. Zhou, and T. Walsh. Fairness amidst non-iid graph data: A literature review. *arXiv preprint arXiv:2202.07170*, 2022.