

Fairer but Not Fair Enough

On the Equitability of Knowledge Tracing

Shayan Doroudi
Carnegie Mellon University Stanford University
Pittsburgh, PA Stanford, CA
shayand@cs.cmu.edu

Emma Brunskill
Stanford University
Stanford, CA
ebrun@cs.stanford.edu

ABSTRACT

Adaptive educational technologies have the capacity to meet the needs of individual students in theory, but in some cases, the degree of personalization might be less than desired, which could lead to inequitable outcomes for students. In this paper, we use simulations to demonstrate that while knowledge tracing algorithms are substantially more equitable than giving all students the same amount of practice, such algorithms can still be inequitable when they rely on inaccurate models. This can arise as a result of two factors: (1) using student models that are fit to aggregate populations of students, and (2) using student models that make incorrect assumptions about student learning. In particular, we demonstrate that both the Bayesian knowledge tracing algorithm and the N -Consecutive Correct Responses heuristic are susceptible to these concerns, but that knowledge tracing with the additive factor model may be more equitable. The broader message of this paper is that when designing learning analytics algorithms, we need to explicitly consider whether the algorithms act fairly with respect to different populations of students, and if not, how we can make our algorithms more equitable.

CCS CONCEPTS

• **Applied computing** → **Computer-managed instruction**; • **Computing methodologies** → *Modeling and simulation*;

KEYWORDS

equity, fairness, knowledge tracing, model misspecification

ACM Reference Format:

Shayan Doroudi and Emma Brunskill. 2019. Fairer but Not Fair Enough On the Equitability of Knowledge Tracing. In *The 9th International Learning Analytics & Knowledge Conference (LAK19)*, March 4–8, 2019, Tempe, AZ, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3303772.3303838>

1 INTRODUCTION

A major challenge in any learning environment is ensuring that students with different needs receive personalized instruction to suit those needs. If instruction is not individualized, then some students may lag behind, while others proceed at a pace that is slower than

ideal. Adaptive educational technology is designed to alleviate some of these concerns by giving students instruction and practice at the pace they need it. While this means that some students may need to work longer than others, ideally all students will eventually be able to master the content at hand. In practice, however, adaptive technologies sometimes fall short of this goal, with lower-performing students receiving less practice and instruction than they need. As some researchers have pointed out, educational technologies that aim to benefit all learners might disproportionately benefit more advantaged groups of learners [6, 12].

In this paper, we examine the equitability of knowledge tracing algorithms that implement mastery learning. We first note that the adaptivity provided by knowledge tracing makes it substantially more equitable than providing all students with the same amount of instruction. However, we show that knowledge tracing algorithms can still be inequitable (favoring fast learners over slow learners) when they rely on inaccurate models of student learning. In particular, we show that this issue could arise in two situations: (1) when using student models that are fit to aggregate populations of students, and (2) when using student models that make incorrect assumptions about student learning.

The first factor was originally observed by Corbett and Anderson [4] when they first introduced knowledge tracing:

The model underestimates the true learning and performance parameters for above-average students. As a result, these students who make few errors receive more remedial exercises than necessary and perform better on the test than expected. In contrast, the model overestimates the true learning and performance parameters for below-average students who make many errors. While these students receive more remedial exercises than the above average students, they nevertheless receive less remedial practice than they need and perform worse on the test than expected.

To correct for this, Corbett and Anderson [4] suggested a way to individualize the parameters of the Bayesian knowledge tracing (BKT) algorithm by using the students' error rates when working on the tutoring system. While their individualization algorithm lead to better predicting student posttest scores, they found in future experiments, which used their individualized BKT algorithm, that lower-performing students still tended to achieve lower posttest scores, even though the model predicted nearly all students had mastered the material. As we show in this paper, this problem could arise when the student model makes incorrect assumptions about student learning. Several other authors have looked at various methods for fitting BKT parameters that are individualized per student

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
LAK19, March 4–8, 2019, Tempe, AZ, USA

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6256-6/19/03...\$15.00
<https://doi.org/10.1145/3303772.3303838>

[9, 10, 15]. Lee and Brunskill [9] in particular looked at the limitations of using a population-level BKT model as opposed to models that are individualized per student and found that a significant percentage of students might be moved to mastery when they haven't in fact mastered the material. However, like Corbett and Anderson [4], the authors did not explicitly consider the equity implications of this or explicitly quantify how inequitable the models could be.

In this paper, we reinterpret these previous findings that using knowledge tracing can lead to different outcomes for high-performing and low-performing students from an equity perspective by taking the view that if an instructional technique leads to higher learning gains for high-performing students than lower-performing students, then that technique is inequitable. We look at two common algorithms used for mastery learning—BKT, which is commonly used in intelligent tutoring systems, and the N -Consecutive Correct Responses (N -CCR) heuristic, which has been used in popular learning platforms such as Khan Academy and ASSISTments [2, 7, 8]—as well as another popular algorithm for modeling student learning, the additive factor model (AFM). We use simulations to examine the behavior of these algorithms when students learn according to either a BKT model or AFM, and find that both BKT and N -CCR could be inequitable. Our findings show that AFM may be a more conservative algorithm under model misspecification, and hence lead to more equitable outcomes. However, our results demonstrate the need for more work in developing robustly equitable knowledge tracing algorithms.

2 VALUE JUDGMENTS

When discussing ethical concerns such as equity, we must necessarily make value judgments. In this section, we define what we mean when we say a knowledge tracing algorithm is more or less equitable. Although we believe our notion of equity is sensible, we do not mean to convince the reader that our definition of equity is the correct definition or the only definition. Rather, we hope that this work generates discussion around when a learning analytics algorithm should or should not be considered equitable.

In this paper, we assume that an equitable outcome is when students from different demographics reach the same level of knowledge after receiving instruction. In what follows, we assume there are only two types of students: slow learners and fast learners. The speed of learning could be thought of as a proxy for separating students who are more or less disadvantaged. For example, slow learners could correspond to students from low socio-economic backgrounds who have had limited access to good instruction or have come into a course with existing gaps in prior knowledge, or slow learners could be coming from a particular demographic background that is more likely to face a stereotype threat or fixed mindset in a particular domain.

Notice that in our notion of equity, the only thing that matters is how much students learned, not how long it took them to reach their level of knowledge. Therefore, we assume it is equitable if slow learners take longer to reach the same level of knowledge as fast learners; it may even be unavoidable. However, if all we cared about was equity, then we could simply give all students inordinate amounts of instruction, but of course that would not be ideal. Therefore, we would ideally want knowledge tracing algorithms

that are equitable while minimizing the amount of time wasted on having students do extra practice opportunities. Moreover, a secondary equity concern could be to have slow learners and fast learners have equal amounts of superfluous practice opportunities. Thus, we will also compare algorithms on how much extra practice they give students, but that is not the primary focus of this paper.

3 KNOWLEDGE TRACING

Knowledge tracing algorithms are used in learning technologies such as intelligent tutoring systems [4, 13], massive open online courses [14], and Khan Academy [7], in order to adaptively assess learners' knowledge states and use that assessment to implement mastery learning (i.e., decide when students have mastered skills and are ready to move on to other skills). We assume that for each skill, students are given a number of practice opportunities for that skill and on each practice opportunity the student will give a response that is either correct or incorrect. The goal of a knowledge tracing algorithm when used for mastery learning is to determine when to stop giving students practice opportunities for the given skill. Knowledge tracing is often performed by a statistical model of student learning that could be fit to data. We will describe two such models and how they can be used to implement mastery learning. In addition to model-based knowledge tracing, simple heuristics can be used to implement mastery learning such as the N -Consecutive Correct Responses (N -CCR) heuristic, which simply gives practice opportunities until the student answers N questions correctly in a row. For simplicity, throughout this paper we assume students are only learning a single skill.

3.1 Bayesian Knowledge Tracing (BKT)

The BKT model is the most commonly used model for knowledge tracing [4, 13, 14]. BKT assumes that for each skill, the student either knows the skill or they do not. Students initially know the skill with probability $P(L_0)$. With every practice opportunity, students have some probability of learning the skill if they do not already know it ($P(T)$). If the student does not know the skill, the student will guess the correct answer with probability $P(G)$ and if the student does know the skill, the student will answer correctly unless they slip with probability $P(S)$. These four parameters fully describe the standard BKT model for each skill. The Bayesian knowledge tracing algorithm proceeds by maintaining a probabilistic belief that the student has learned the skill given the sequence of observed responses and the parameters of the BKT model. Once this probability exceeds some mastery threshold (typically taken to be 0.95), the algorithm decides that the student has mastered the skill.

One can learn the parameters of the BKT model by fitting it to a dataset consisting of sequences of student responses. In all of the simulations below, we fit the BKT models using brute-force grid search over the entire parameter space in 0.01 increments with the BKT Brute Force model fitting code [1].

3.2 Additive Factor Model (AFM)

AFM is another popular model of student learning, but it is not typically used to perform knowledge tracing; rather, AFM is often used to examine learning curves after students go through mastery learning [3]. Unlike BKT, AFM assumes that learning takes place incrementally; with each practice opportunity, the probability that

	BKT-Slow	BKT-Fast	BKT-Mixed
$P(L_0)$	0.0	0.0	0.071
$P(T)$	0.05	0.3	0.096
$P(G)$	0.2	0.2	0.209
$P(S)$	0.2	0.2	0.203

Table 1: BKT Models used in simulations in Section 4.

the student will answer correctly on future practice opportunities increases. This also means that unlike BKT, AFM does not try to predict a latent knowledge state (whether the student has learned the skill or not), but rather directly models student performance (the probability of answering correctly at any given time). In particular, a simplified version of AFM for when there is only one skill is governed by the following logistic function:

$$P(C_i = 1) = \frac{1}{1 + \exp(-(\theta - \beta + \gamma i))}$$

where $P(C_i = 1)$ is the probability that the student will answer the i th practice opportunity correctly, θ is the student ability (which could encapsulate the student’s prior knowledge), β is the difficulty of the skill, and γ is the learning rate. Since we are only interested in a single skill, we set $\beta = 0$ and let θ combine the student ability and the item difficulty. Notice that another difference between AFM and BKT is that AFM allows for individual differences via θ . Thus, if we want to use AFM for knowledge tracing, assuming we have not interacted with a particular student before, we would need to estimate θ online (as is done in computerized adaptive testing). When using AFM to implement mastery learning, we need to choose a certain level of desired accuracy for the mastery threshold.

4 CASE 1: LACK OF INDIVIDUALIZATION

We first demonstrate what happens when the BKT mastery learning algorithm uses a model that is fit to data coming from a mix of slow learners and fast learners. The BKT models we used are shown in Table 1. Notice that the two models, BKT-Slow and BKT-Fast, only differ in the $P(T)$ parameter, which we will refer to as the learning rate. Table 1 also shows BKT-Mixed, the model that was fit to 200 simulated students from a population of students who were equally likely to come from BKT-Slow and BKT-Fast and who received 20 practice opportunities each. Notice that the key difference between BKT-Mixed and the other models is that it has an intermediate learning rate, since the best fitting model is trying to average over the different rates of learning that were present in the data.

To assess whether the BKT-Mixed algorithm would behave equitably, we ran simulations where BKT-Mixed was used to implement mastery learning for both slow and fast learners. As mentioned before, we are primarily interested in comparing the percentage of slow learners and fast learners who did not learn the skill (even though the algorithm believes they did with 95% certainty), but as a secondary concern, we are also interested in the average amount of extra practice for students who did learn the skill. We first compare the BKT-Mixed algorithm to non-adaptive instructional policies that give a fixed amount of practice to all students. Figure 1 shows that for non-adaptive instructional policies, there is a tradeoff between the equity gap between slow and fast learners and the amount of extra practice given. However, BKT-Mixed achieves a much better

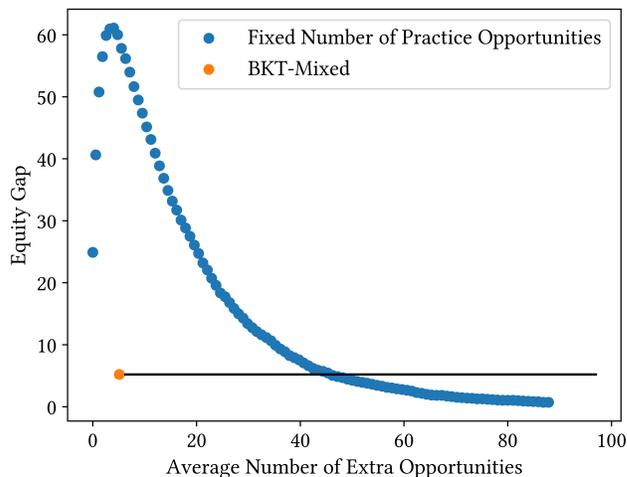


Figure 1: Comparison of mastery learning using BKT-Mixed with non-adaptive policies that give a fixed number of practice opportunities (between 1 and 100) to all students. The equity gap refers to the % of slow learners - % of fast learners that do not learn the skill. The x-axis shows the average amount of extra practice for students that learned the skill.

balance between equity and extra practice than the non-adaptive policies. In particular, to achieve the same gap between slow and fast learners for a non-adaptive policy, one would have to give around 46 extra practice opportunities to students on average.

However, Figure 1 also confirms that the BKT-Mixed algorithm is inequitable, even if much less than non-adaptive algorithms. As shown in Table 2, the BKT-Mixed algorithm leads to more slow students (5.5%) not learning the skill than fast students (0.3%), while giving roughly the same amount of extra practice for both slow and fast learners. This problem could be resolved in one of two ways. First, individualized BKT models could be fit per student or per cluster of students to account for the different learning rates. Second, the mastery threshold could be increased, at the expense of giving learners more unnecessary practice. The second row of Table 2 shows that if we had used the true model (BKT-Slow for slow learners and BKT-Fast for fast learners), the gap between slow and fast learners diminishes¹. Notice that using these individualized models, the outcomes for fast learners are seemingly worse than when using BKT-Mixed, given our stance that not learning the skill is worse than receiving additional practice. This is simply because we are using a threshold of 95%, which states that we are comfortable with 5% of students not reaching mastery; if we would rather have more students reach mastery at the expense of students receiving extra practice we could simply increase the mastery threshold. In some sense, the mastery threshold represents the degree to which we are willing to tradeoff between the percentage of students that learn the skill and the amount of extra practice.

In addition to the BKT algorithm, Table 2 also shows that the the 3-CCR heuristic, which has been used in the ASSISTments system

¹There are still 1% more slow learners who do not reach mastery than fast learners, but this difference is not systematic; for some learning rates, slow learners have a slightly higher percentage of learning the skill instead.

	Slow Learners			Fast Learners		
	% Not Learned	Avg Extra	Avg % Extra	% Not Learned	Avg Extra	Avg % Extra
BKT-Mixed	5.5%	5.0	89%	0.3%	5.1	269%
True Models	3.0%	5.6	96.8%	2.0%	3.3	180%
3-CCR	10.3%	4.4	82%	0.9%	4.4	230%
5-CCR	0.4%	9.9	159%	0.02%	9.8	497%
AFM	0.2%	19.4	193%	0.02%	9.7	472%

Table 2: Effects of simulating various mastery algorithms on simulated BKT learners with different learning rates. Each simulation consisted of either 5000 BKT-Slow learners or 5000 BKT-Fast learners who were given practice opportunities until they reached mastery according to the knowledge tracing algorithm for that row. For each type of learner, the first column shows the percentage of students that did not learn the skill, the second column shows the average number of superfluous practice opportunities for students who learned the skill, and the third column shows the average percentage increase in number of practice opportunities beyond those that the student needed to learn the skill.

[2], is also inequitable. In fact, the achievement gap is higher than for BKT (10.3% of slow learners do not learn the skill vs. 0.9% of fast learners). Notice that although N -CCR does not explicitly use a model, it still makes the assumption that all learners are the same, which is the root of the problem. Again, this could be fixed either by choosing an individualized value of N for each student or each type of learner (e.g., slow and fast), or by uniformly increasing the value of N , which is analogous to increasing the mastery threshold for BKT. In this case, $N = 5$ is sufficient, as seen in Table 2. However, it may be unclear a priori how large N needs to be, and if N is too large, students may be doing much more practice than necessary. For example, Khan Academy found that even with $N = 10$, students who took more problems to get 10 correct in a row (presumably slower learners) were more likely to get the next question incorrect than students that immediately got 10 in a row correct [7].

5 CASE 2: MODEL MISSPECIFICATION

In the previous section, we saw that if there are students that learn at different rates, and we do not account for that in modeling student learning, then knowledge tracing could lead to widening the achievement gap. In this section, we look at the more fundamental problem where the model of student learning is entirely misspecified. In particular, we show that if learning is described by the additive factor model, then using BKT can lead to inequitable outcomes, *even if* we fit different parameters for fast and slow learners. We defined two AFM models with the only difference being that one model had a learning rate (γ) of 0.05 (AFM-Slow) and the other model had a learning rate of 0.1 (AFM-Fast). For both models, the student abilities (θ) were sampled from a $\text{NORMAL}(-2, 1)$ distribution. We then fit a BKT model to 200 simulated students from a population of students who were equally likely to come from AFM-Slow and AFM-Fast and who received 20 practice opportunities each. We refer to this model as $\text{BKT-Mixed}_{\text{AFM}}$. We also fit a BKT model to a population of only 200 AFM-Slow learners ($\text{BKT-Slow}_{\text{AFM}}$) and a BKT model to a population of only 200 AFM-Fast learners ($\text{BKT-Fast}_{\text{AFM}}$). These models are shown in Table 3. Because we are trying to fit data where learning happens incrementally with a model where learning is all-or-nothing, the slip parameters are abnormally large; prior work has discussed this phenomenon and how these parameters cannot be given the normal interpretation of

	$\text{BKT-Slow}_{\text{AFM}}$	$\text{BKT-Fast}_{\text{AFM}}$	$\text{BKT-Mixed}_{\text{AFM}}$
$P(L_0)$	0.123	0.055	0.065
$P(T)$	0.021	0.043	0.022
$P(G)$	0.123	0.151	0.152
$P(S)$	0.448	0.435	0.41

Table 3: BKT Models used in simulations where AFM is the true model that describes student learning.

	Slow Learners	Fast Learners
$\text{BKT-Mixed}_{\text{AFM}}$	0.49	0.60
$\text{BKT-Slow}_{\text{AFM}}$	0.45	-
$\text{BKT-Fast}_{\text{AFM}}$	-	0.56
3-CCR	0.47	0.58
5-CCR	0.67	0.76
AFM	0.955	0.962

Table 4: Effects of simulating various mastery learning algorithms on simulated AFM learners with different learning rates. Each simulation consisted of either 500 AFM-Slow learners or 500 AFM-Fast learners who were given practice opportunities until they reached mastery according to the knowledge tracing algorithm for that row. Both columns show the average probability of answering an item correctly when the knowledge tracing algorithm declares mastery.

slipping [4, 5, 11]. Moreover, notice that the key difference between the models fit to slow and fast learners is the learning rate, which makes sense as that is the only difference between the AFM models; interestingly, the mixed model has a learning rate that is almost identical to that of the slow model, but the rest of the parameters have values closer to those of the fast model.

Because AFM is an incremental model of learning, it no longer makes sense to evaluate whether or not students learned the skill or how many extra practice opportunities they received. Instead, we are interested in the probability of students answering the next item correctly at the point where the knowledge tracing algorithm declares mastery. An equitable outcome would be to have an identical probability of correct answers for fast and slow learners. Table 4 shows the results of simulations of knowledge tracing with the three

fitted BKT models on AFM-Slow students and AFM-Fast students. What we find is that there is a 0.11 difference in the probability of correctness at mastery for slow and fast learners. Moreover, this difference does not go away even when we individualize the BKT models (i.e., use the BKT models specifically fit to slow and fast learners). A similar gap exists when using the 3-CCR algorithm, and in this case the gap barely decreases when using the 5-CCR algorithm. Thus, when the model used in knowledge tracing is incorrect, even individualization of the model may not be sufficient to eliminate the achievement gap.

To demonstrate that the problem in this case is indeed model misspecification, we also used AFM to perform knowledge tracing. Recall that AFM assumes each student has an individualized ability parameter (θ). Assuming we do not have knowledge of this ability ahead of time, it makes sense to estimate it in an online fashion. Moreover, because we also want to distinguish between slow and fast learners, it would be ideal if we could learn the learning rate (γ) online. Thus, to use AFM for knowledge tracing, we fit the logistic regression model at each time step using all the practice opportunities for that student so far as data. Additionally, we decided that the AFM-based algorithm would declare mastery when the estimated probability of correctness reached 0.95. To ensure that the student had actually reached the threshold (rather than the parameters being spuriously estimated), we used the minimum estimated probability of correctness over the last three practice opportunities. As shown in Table 4, this algorithm was equitable in that slow and fast learners both reached the desired level of mastery (0.95).

Since this online AFM-based knowledge tracing can estimate the learning rate on-the-fly, we wanted to see if it would be equitable when used on BKT learners. To test the AFM algorithm on BKT learners, we used the same procedure as described above, except that we set the mastery threshold to 0.8, as that is the highest accuracy that students can reach under a BKT model with a slip probability of 0.2. The last row of Table 2 shows that using this algorithm leads to mastery for all learners, at the expense of giving more practice opportunities than needed, especially for slow learners. While the amount of extra practice for slow learners is not ideal, this suggests that perhaps the AFM algorithm is more equitable, even under model misspecification. However, more work is needed in assessing the equitability of AFM-based knowledge tracing assuming different types of model misspecification.

6 CONCLUSION AND FUTURE DIRECTIONS

By allowing all students to go through curricula at their own pace to ultimately reach mastery, adaptive educational algorithms such as knowledge tracing are implicitly meant to eliminate inequities between different groups of students. While this is a noble goal, we have shown that due to differences between the way we model student learning and how students actually learn, we may sometimes fall short of reaching this goal. Although the degree of inequity in some cases may seem small (e.g., 5% of slow learners are lagged behind), these inequities could cascade as the amount of content increases and as content builds on assumed prior knowledge. However, the observation that using AFM might lead to more equitable outcomes makes us optimistic that there exist approaches to knowledge tracing that are equitable in a variety of settings.

We believe there are several concrete steps that can be taken to design more equitable knowledge tracing algorithms. First, we should consider the equity of knowledge tracing in more realistic settings, such as when students are learning multiple skills sequentially. A proper notion of equity in this setting must consider both the degree to which skills are learned as well as how many skills are learned. Second, we should evaluate the equitability of different knowledge tracing algorithms under various assumptions about how students learn. This could lead to finding ways to refine our algorithms to make them more equitable in diverse settings. Finally, we should move beyond simulations to see if algorithms that are predicted to be equitable actually live up to that promise when used to teach actual learners, and if not, how we can refine our simulations to develop more equitable algorithms. Our hope is that this paper encourages designers of learning analytics algorithms (beyond just knowledge tracing) to explicitly consider equity when designing their algorithms.

ACKNOWLEDGEMENTS

This research was supported in part by a Google grant and a Microsoft Research Faculty Fellowship.

REFERENCES

- [1] Ryan S Baker, Albert T Corbett, Sujith M Gowda, Angela Z Wagner, Benjamin A MacLaren, Linda R Kauffman, Aaron P Mitchell, and Stephen Giguere. 2010. Contextual slip and prediction of student performance after use of an intelligent tutor. In *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization*. Springer, 52–63.
- [2] Joseph E Beck and Yue Gong. 2013. Wheel-spinning: Students who fail to master a skill. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*. Springer, 431–440.
- [3] Hao Cen. 2009. *Generalized learning factors analysis: Improving cognitive models with machine learning*. Ph.D. Dissertation.
- [4] Albert T Corbett and John R Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 4 (1994), 253–278.
- [5] Shayan Doroudi and Emma Brunskill. 2017. The misidentified identifiability problem of Bayesian Knowledge Tracing. In *Proceedings of the 10th International Conference on Educational Data Mining*, 143–149.
- [6] John D Hansen and Justin Reich. 2015. Democratizing education? Examining access and usage patterns in massive open online courses. *Science* 350, 6265 (2015), 1245–1248.
- [7] David Hu. 2011. How Khan Academy is using machine learning to assess student mastery. <http://david-hu.com/2011/11/02/how-khan-academy-is-using-machine-learning-to-assess-student-mastery.html>
- [8] Kim M. Kelly, Yan Wang, Tamisha Thompson, and Neil T. Heffernan. 2015. Defining mastery: Knowledge tracing versus N-consecutive correct responses. In *Proceedings of the 8th International Conference on Educational Data Mining*.
- [9] Jung In Lee and Emma Brunskill. 2012. The impact on individualizing student models on necessary practice opportunities. In *Proceedings of the 5th International Conference on Educational Data Mining*.
- [10] Zachary A Pardos and Neil T Heffernan. 2010. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 255–266.
- [11] Radek Pelánek. 2018. Conceptual issues in mastery criteria: Differentiating uncertainty and degrees of knowledge. In *Proceedings of the 19th International Conference on Artificial Intelligence in Education*. Springer, 450–461.
- [12] Justin Reich and Mizuko Ito. 2017. *From good intentions to real outcomes: Equity by design in learning technologies*. Digital Media and Learning Research Hub.
- [13] Steven Ritter, John R Anderson, Kenneth R Koedinger, and Albert Corbett. 2007. Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review* 14, 2 (2007), 249–255.
- [14] Yigal Rosen, Iliia Rushkin, Rob Rubin, Liberty Munson, Andrew Ang, Gregory Weber, Glenn Lopez, and Dustin Tingley. 2018. The effects of adaptive learning in a massive open online course on learners’ skill development. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. ACM, 6.
- [15] Michael V Yudelson, Kenneth R Koedinger, and Geoffrey J Gordon. 2013. Individualized bayesian knowledge tracing models. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*. Springer, 171–180.