PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

An astonishing regularity in student learning rate

Kenneth R. Koedinger[a,1], Paulo F. Carvalho[a], Ran Liu[b], and Elizabeth A. McLaughlin[a]

Edited by Douglas Medin, Northwestern University, Evanston, IL; received December 25, 2022; accepted February 10, 2023

Leveraging a scientific infrastructure for exploring how students learn, we have developed cognitive and statistical models of skill acquisition and used them to understand fundamental similarities and differences across learners. Our primary question was why do some students learn faster than others? Or, do they? We model data from student performance on groups of tasks that assess the same skill component and that provide follow-up instruction on student errors. Our models estimate, for both students and skills, initial correctness and learning rate, that is, the increase in correctness after each practice opportunity. We applied our models to 1.3 million observations across 27 datasets of student interactions with online practice systems in the context of elementary to college courses in math, science, and language. Despite the availability of up-front verbal instruction, like lectures and readings, students demonstrate modest initial prepractice performance, at about 65% accuracy. Despite being in the same course, students' initial performance varies substantially from about 55% correct for those in the lower half to 75% for those in the upper half. In contrast, and much to our surprise, we found students to be astonishingly similar in estimated learning rate, typically increasing by about 0.1 log odds or 2.5% in accuracy per opportunity. These findings pose a challenge for theories of learning to explain the odd combination of large variation in student initial performance and striking regularity in student learning rate.

learning rate | learning curves | deliberate practice | logistic regression growth modeling; educational equity

Humans are capable of a wide and flexible variety of learning adaptation. This adaptability is particularly apparent in the development of expertise associated with high-profile careers, like technology innovation or music composition, but also in the wide variety of academic subject matter, reading, writing, math, science, second language, etc., humans master. Better understanding of how human learning works in the context of academic courses is of scientific interest because academic learning is particularly distinct to the human species. It is also of practical interest because such understanding can be used to develop more effective education. New technologies have often made better science possible. Such is the case for educational technologies which, in this century, have been increasingly providing unprecedented volumes of detailed data on academic learning. With center-level funding from the National Science Foundation to LearnLab (learnlab.org), we developed a social–technical infrastructure to systematically acquire such data and use it both to optimize interactive learning technologies and to pursue scientific questions about student learning.

LearnLab's early goals were to identify the mental units of learning in academic courses, to use these insights to design and demonstrate improved instruction in randomized controlled experiments embedded in courses, and to build models of learners that may reveal significant similarities and differences across learners. Past research produced methods for discovering and validating improved cognitive models of the mental units students acquire in academic courses (e.g., ref. 1). These improved cognitive models were used to redesign course units, and random assignment field experiments comparing student use of the redesign (treatment) with the original design (control) demonstrated enhanced learning outcomes (e.g., refs. 2 and 3). A key theoretical hypothesis of these cognitive models is that a decomposition of learning into discrete units, or knowledge components, produces predictions that can be tested against student performance data across different contexts and at different times. Investigations across multiple datasets support this knowledge component hypothesis (e.g., refs. 1 and 4).

In this paper, we combine these cognitive models with statistical growth models to explore significant similarities and differences across academic learners. Our research questions are:

1. Practice needed: How many practice opportunities do students need to reach a mastery level of 80% correctness?
2. Initial performance variation: How much do students vary in their initial performance?
3. Learning-rate variation: How much do students vary in their learning rate?

Significance

Prior research, often using self-report data, hypothesizes that the path to expertise requires extensive practice and that different learners acquire competence at different rates. Fitting cognitive and statistical growth models to 27 datasets involving observations of learning and performance in academic settings, we find evidence for the first hypothesis and against the second. Students do need extensive practice, about seven opportunities per component of knowledge. Students do not show substantial differences in their rate of learning. These results provide a challenge for learning theory to explain this striking similarity in student learning rate. They also suggest that educational achievement gaps come from differences in learning opportunities and that better access to such opportunities can help close those gaps.

Author affiliations: [a]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213; and [b]Engineering, Amira Learning, Seattle, WA 98101

Author contributions: K.R.K., P.F.C, and R.L. designed research; K.R.K., P.F.C., and R.L. performed research; K.R.K., P.F.C., and R.L. analyzed data; and K.R.K., P.F.C., and E.A.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: koedinger@cmu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221311120/-/DCSupplemental.

Published March 20, 2023.

Research question 1 probes how much practice, if any, students need beyond the up-front verbal instruction they typically receive (e.g., from course lectures and/or reading assignments) before practice begins. We find that students are not at mastery at the start of practice, and substantial learning occurs from the practice itself as students receive feedback on their performance and make use of context-sensitive verbal instruction and examples. We find that a typical student needs about seven learning opportunities to master a typical knowledge component. However, we find substantial variation in needed opportunities across students. Thus, questions 2 and 3 probe whether that variation is more due to differences in initial performance and/or differences in the rate at which performance improves with each successive learning opportunity.

A rigorous and broad estimation of variation in student learning rate informs important scientific debates. Research on expertise (5) indicates that even historic geniuses needed years of practice to develop their expertise. Ericsson (6) estimates that high-level expertise takes about 10,000 h of practice to develop and claims that no substantial exceptions have been found. In other words, no matter who you are, you need many repeated practice opportunities to develop expertise. As a counterpoint, other researchers (e.g., refs. 7 and 8) have suggested that practice time alone is not enough to account for how much expertise is acquired and that some people may learn more (or less) than others given the same practice time. This debate comes down to whether learning rate per practice opportunity is relatively constant across individuals or whether it varies substantially. It has been difficult to make progress in resolving or refining this debate because of limitations in available data. Existing data fueling this debate come from a small number of participants, are prone to subjective error as they are largely self-reported, and are sparse and coarse grained (few data points per participant). In contrast, the datasets we have accumulated directly track practice and are thus objective and are fine grained (about 200 observations per student), lasting over hours or weeks, and are large in students (nearly 7,000 students).

Beyond the deliberate practice debate, we find some researchers indicating substantial differences in student learning rate (9, 10) and others indicating little difference in student learning rate (11). Consider, for instance, a National Academy of Sciences report indicating that “high-ability learners learn at a more rapid rate than other students” (9, p. 131). In contrast to the National Academy of Sciences report, Bloom (11) suggested that “most students become very similar with regard to … rate of learning … when provided with favorable learning conditions.” (p. *x*). While Bloom and colleagues did demonstrate the effectiveness of a form of deliberate practice, they did not provide evidence to demonstrate their claim of uniformity in learning rate. Nor does the National Academy of Sciences report point to evidence for learning-rate variability. This project provides an opportunity to test these competing claims.

Importantly, the claim in the National Academy of Sciences report is about high-ability learners, suggesting differences due to learner characteristics. It can be contrasted with a claim, which is almost certainly true, that learners in more favorable conditions learn at a more rapid rate than those in less favorable conditions. The educational technologies used in our NSF-funded LearnLab studies arguably provide favorable learning conditions as they implement research-based principles (e.g., varied practice with feedback and explanatory instruction), and many have been improved through iterative data-driven cognitive task analysis and experimental methods (12). A key goal of LearnLab was to identify, in the words of the National Academy of Sciences report, high-ability learners who “learn at a more rapid rate than other students” (13, p. 37). We were interested in identifying differences in students’ self-regulated learning skills or background knowledge that would yield learning-rate differences and that might be addressed through instructional support for learning to learn. Thus, we were quite surprised as results began to emerge suggesting an astonishing amount of regularity in student learning rate (14).

One may be tempted by everyday experience to suggest there is obvious wide variability in how fast different people learn. At the end of an algebra course, for example, some students are getting an A and appear to have learned faster than other students who are getting low grades. However, such differences may be alternatively explained not as differences in learning rate but as differences in the number of quality learning opportunities individuals experience. In the varied data sources we have accumulated, the number of learning opportunities students experience is known, and thus we can gain insight into whether student competence differences derive more from environmental opportunity differences or student-inherent learning-rate differences.

In particular, we model learning using 27 datasets with over 1.3 million student performance observations from 6,946 learners in 12 different courses ranging across math, science, and language learning, across educational levels from late elementary to college, and across educational technologies including intelligent tutoring systems, educational games, and online courses (*SI Appendix*, Table S1).

Should student performance captured in these datasets be considered representative of human learning generally? These datasets were produced by students using educational technology in natural contexts of academic courses. These courses involved common forms of instruction, such as lectures and assigned readings, which typically preceded student practice within the educational technology. While student practice has historically been done mostly on paper, we suspect, as in modern psychological experiments where participants interact with technology, that the technology interaction itself is not substantially changing the psychological processes involved. Thus, these datasets are arguably well representative of complex human learning as it is displayed in academic contexts in math, science, and language learning.

Should these educational technology contexts be considered “favorable learning conditions” per Bloom’s claim? These contexts are prime examples of a learning-by-doing approach that has been repeatedly advocated in different variations and with substantial experimental support, including “mastery-based learning” (15), “active learning” (16), “testing effect” (17), “formative assessment” (18), and “deliberate practice” (5, 19, 20). These contexts provide favorable learning conditions not only because of the active learning support, but also because of more particular features of learning interactions each with its own scientific basis. All these educational technologies a) provide immediate feedback on errors in problem solving or performance contexts (21, 22), b) provide explanatory context-specific instruction on demand (e.g., ref. 23), including an example correct response if needed (24–26), c) highly encourage or enforce students to enter or observe a correct response before moving on, d) provide tailored tasks designed through data-based cognitive task analysis to practice specific cognitive competences aligned with course goals for improving student thinking (e.g., refs. 27 and 28), and e) give repeated opportunities to ensure student mastery of these cognitive competences (e.g., ref. 29) in varied tasks that require appropriate generalized, but not overgeneralized, knowledge and skill acquisition (e.g., ref. 30).

Modeling Learning by Integrating a Cognitive Model into a Logistic Regression Growth Model

To model student performance and learning, we used mixed effects logistic regression with a cognitive model component and a growth component. As indicated in the first line in Fig. 1, we model the

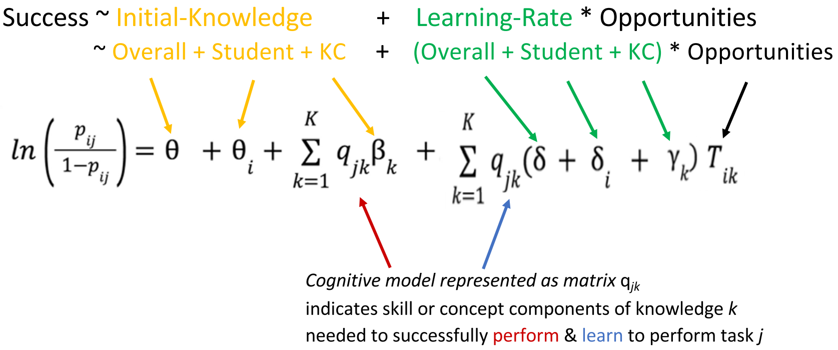
Success ~ Initial-Knowledge + Learning-Rate * Opportunities
~ Overall + Student + KC + (Overall + Student + KC) * Opportunities
$ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \theta + \theta_i + \sum_{k=1}^{K} q_{jk}\beta_k + \sum_{k=1}^{K} q_{jk}(\delta + \delta_i + \gamma_k)\, T_{ik}$
Cognitive model represented as matrix q_{jk}
indicates skill or concept components of knowledge k
needed to successfully perform & learn to perform task j

Fig. 1. We model success (p_{ij}) of student i on task j across deliberate practice opportunities (T_{ik}) in a logistic regression with initial-knowledge and learning-rate estimates. These estimates are each decomposed into overall, student (i), and knowledge component (k) elements. The knowledge components required by each task (j) are specified in a cognitive model matrix (q_{jk}).

success of student i on task j as proportional to a linear function with an intercept that represents initial-knowledge (shown in orange) and a slope that represents learning-rate (shown in green) per opportunity to learn. Both initial-knowledge and learning-rate are further broken down (second line in Fig. 1) to provide an overall estimate and variations due to student and knowledge component (KC). These factors lead to six parameters (predictor variables) in a mixed effects logistic regression (right side of equation in Fig. 1) where the outcome variable (p_{ij}) is the probability that student i gets task j correct (left side but shown transformed by the log odds as per logistic regression). The six predictor variables include two fixed effects for overall initial-knowledge (θ) and overall learning rate (δ) and four random effects for the student initial-knowledge (θ_i), student learning-rate (δ_i), KC initial-knowledge (β_k), and KC learning-rate (γ_k) (see *SI Appendix* for the precise model description in R). The three learning-rate parameters (in green) are multiplied by the number of opportunities (T_{ik}) student i has experienced on knowledge component k.

The product of learning rates by opportunity is the growth component of this model. Another key component is the cognitive model, which is represented by matrix q_{jk}. This matrix indicates for each task j what knowledge component is needed to perform that task[*]. In general, a cognitive model is an explanation of one or more cognitive processes that generates predictions that can be matched to human data (31). These cognitive models are implemented as computational models in datasets coming from Intelligent Tutoring Systems (see “ITS” in the Ed Tech column of *SI Appendix*, Table S1). The q_{jk} matrix provides a simplified representation of a cognitive model useful for statistical analysis. Many prior investigations have evaluated and refined these cognitive or knowledge component (KC) models by comparing alternative versions of the q_{jk} matrix (27, 28, 32–35). This KC model refinement process is illustrated in Table 1 with four different q_{jk} matrices for the same six tasks. As noted above, a task observation is often a step in a problem solution such that, for example, the problem 2 * 8 – 30 is observed in two steps as shown in the first two rows of Table 1.

Each alternative q_{jk} introduces different hypotheses about what makes tasks difficult (see the β_k term in Fig. 1) and what yields transfer of learning across practice opportunities on related tasks (the term γ_k). Q0 reflects the hypothesis that all arithmetic tasks require one knowledge component (KC) and predicts all tasks will be of similar difficulty and practice on any one improves performance on another. Q1 separates multiplication and subtraction as different KCs. One may further hypothesize, as in Q2, that learning to solve tasks like the third (30 – 2 * 8 → 30 – 16) and the fifth requires extra order of operations knowledge (MultOR), whereas the first task (2 * 8 – 30 → 16 – 30) can be solved by a naive left-to-right strategy (MultLR). Q3 represents hypotheses that each task has its own inherent difficulty and that there is no transfer of learning across tasks. When q_{ik} is the identity matrix, as in Q3, the initial knowledge terms in Fig. 1 are equivalent to item response theory (36). Our model is similar to others who have used generalizations of item response theory to model student response data (e.g., refs. 37 and 38).

A KC model can be selected by comparing which q_{jk} matrix provides the best prediction fit to the student data. Care must be taken to use fitness measures that prevent overfitting due to increasing complexity (Q0-Q4) either by penalizing for greater parameters (e.g., using the Akaike Information Criterion, AIC, or the Bayesian Information Criterion, BIC) or testing on held-out data via cross validation. DataShop facilitates such comparisons using a simpler version of the model in Fig. 1 where the student learning-rate term (δ_i) is not included. We selected the best KC model for each dataset as discussed in *Materials and Methods*.

Results

Illustrating and Evaluating a Learning Growth Model. Fig. 2*A* shows a learning curve (in gold) from one of our datasets (ds394 in *SI Appendix*, Table S1) where overall average probability correct (p_{ij}), on the y axis, is increasing with successive opportunities (T_{ik}), on the x axis. The model predictions are shown in green. Individual student curves derived from the model are shown in Fig. 2*B*. It is difficult to visually compare student learning rates (e.g., is student S1 faster than S2?) in this nonlinear scale where the opportunity-to-opportunity increase (e.g., 2.9% from opportunity 1 to 2 for S3) gets smaller at higher opportunities (2.0% from 6 to 7). Rescaling success using log odds, as shown in Fig. 2*C*, produces a linear relationship whereby nuanced differences in student learning rate are apparent (e.g., S2 is steeper than S1 and S3). Fig. 2*D* shows initial knowledge estimates (the intercepts in yellow) and learning-rate estimates (the slopes in blue) for all three students, both in log odds.

We performed comparisons to evaluate the explanatory value of including student learning rate and measuring learning rate in terms of discrete opportunities rather than time. Much of the prior research in refining KC-based cognitive models (1, 39) has used a simpler version of the model in Fig. 1 where the student learning-rate term (δ_i) is not included. This simpler model has come to be called the Additive Factors Model (AFM)[†]. The model in Fig. 1 that includes individual student learning-rate parameters (δ_i) is called individual AFM or iAFM (14). If iAFM provides a good

[*] The general statistical model allows for a task to be labeled by multiple KCs but such models tend not to produce better predictions than single KC models, and the single KC models offer greater simplicity of interpretation.

[†] The factors are “additive” because of the summation of log-odds values associated with KCs or “factors” in task difficulty and transfer. This addition of log-odds is in contrast with the multiplication of probabilities associated with KC factors in a more complicated Conjunctive Factors Model (39).

Table 1. Four alternative cognitive models for the same tasks represented as q_{jk} KC models

	Q0	Q1		Q2				Q3 = Item model					
Tasks *j* (Observed problem steps)	Arith	Mult	Sub	MultLR	MultOR	Sub+	Sub−	I1	I2	I3	I4	I5	I6
2*8−30→ 16−30	1	1	0	1	0	0	0	1	0	0	0	0	0
16−30→ −14	1	0	1	0	0	0	1	0	1	0	0	0	0
30−2*8→ 30−16	1	1	0	0	1	0	0	0	0	1	0	0	0
30−16→ 14	1	0	1	0	0	1	0	0	0	0	1	0	0
10−3*7→ 10−21	1	1	0	0	1	0	0	0	0	0	0	1	0
10−21→ −11	1	0	1	0	0	0	1	0	0	0	0	0	1

Note: Column groups Q0, Q1, Q2, and Q3 represent different knowledge component (KC) models. Each row represents a task. 1 indicates that, in a given knowledge component model, the column KC is required in that task (row). For example, for Q0, all tasks require arithmetic (Arith), whereas for Q2, tasks are hypothesized to require four different KCs depending on the operation involved and the context of its use.

model for detecting student learning-rate variation, we should see that it outperforms AFM, at least on a substantial number of datasets. To be sure, such better performance of iAFM is neutral regarding the size of the student learning-rate variation—a topic we explicitly address further below.

AFM was the best fitting model in six out of 27 datasets based on the Akaike Information Criterion (AIC) and 12 out of 27 datasets based on the Bayesian Information Criterion (BIC), whereas iAFM was the best fitting model in 21 out of 27 datasets (AIC based) or 15 out of 27 datasets (BIC based). We used Wagenmakers (40) approximation to derive Bayes factors from BIC. Across the 27 datasets, 15 had strong or greater evidence in favor of iAFM being the generating model (Bayes Factor > 1), whereas 11 had strong or greater evidence in favor of AFM being the generating model (Bayes Factor < −1). These results indicate that student learning-rate variation is present and detectable in some datasets. In other datasets, it is either not present or too small to be detectable. As we elaborate later, even when student learning-rate variation is detectable, it is not particularly large.

To evaluate the hypothesis that students learn as a result of KC-specific practice opportunities within the educational technologies, we contrast an additional model. The time-based Additive Factors Model (Time-AFM) implements the alternative hypothesis that students learn from general accumulated experiences in and outside the technology using the elapsed calendar time to predict each performance observation. In this model, we replaced the count of practice opportunities used in iAFM and AFM (T_{ik}) with a calendar time variable. That is, to predict performance (p_{ij}) of student *i* on task *j*, we use the calendar time C_{ik} that has passed since this student first experienced knowledge component *k* associated with this task *j*. In Time-AFM, C_{ik} takes the place of T_{ik} in AFM. The results indicate that, despite its considerable overlap with AFM and iAFM, Time-Based AFM was only the best fitting model in one out of the 20 datasets (seven datasets did not have appropriate time-logging to run this comparison). Thus, we have clear evidence that learning growth is better characterized by KC-specific practice opportunities within the technology than by a calendar time variable that also reflects general opportunities for growth and for out-of-technology learning.

Students Start at About 65% Correctness and Need about 7 Practice Repetitions. Using the iAFM modeling results, we investigated students' typical initial performance. This investigation has relevance to the question of whether initial verbal instruction, most typically in the form of readings and lectures, is sufficient for reaching a reasonable level of mastery (defined as 80% correctness) (15). One possibility is that verbal instruction is enough for the average student to reach mastery and deliberate practice will just serve to strengthen what is learned, making performance faster and more fluent. Conversely, it is possible that verbal instruction is not sufficient to reach reasonable accuracy and deliberate practice is needed. As a measure of initial performance, we used the population iAFM intercept for each dataset (see θ in Fig. 1). *SI Appendix*, Table S1 shows the population intercept (initial knowledge estimate) for each dataset. The median population intercept (θ) across the 27 datasets is 0.638 log odds (M = 0.501, 95% CI = [0.285,0.718]), which converts to 65.42% correct (M = 61.79%, 95% CI = [56.77%, 66.82%]). That is, the typical student is starting practice well below mastery despite having been provided with verbal instruction in the form of readings and lectures prior to the deliberate practice experiences in our data.

Having established that with verbal instruction alone, students did not reach mastery, we investigated how much deliberate practice students tend to need. The median overall learning rate (δ) across datasets is 0.09 log odds (M = 0.15, 95% CI = [0.08, 0.22]) per opportunity, which converts to a 2.5 percentage point increase for one practice opportunity from the median intercept (θ) of 65%. We used the overall parameter estimates (θ and δ) from each dataset [with this formula (log-odds (0.80) – log-odds(θ))/δ] to determine how many opportunities a typical student from that dataset would need to reach mastery. Across all datasets, the median number of opportunities to reach mastery is 7.24 (M = 12.27, 95% CI = [7.09, 17.45]). In other words, a typical student learning a typical KC tends to require seven additional practice opportunities to reach mastery after noninteractive verbal instruction (i.e., text or lecture).

Students Vary Substantially in Initial Knowledge. We investigated how much students vary in their initial knowledge using the model fits from iAFM. For each dataset, we computed the SD of student intercepts ($\theta + \theta_i$) and found the median standard across datasets to be 0.651 (M = 0.724, SD = 0.283) and the median interquartile range is 0.830 (M = 0.988, SD = 0.430) in log odds[‡]. This large variation is more apparent if we compare the median student intercept of the lower and upper halves of student intercepts. When converted to percentages, we see (first column of Table 2) that students in the lower half of initial knowledge had a median correctness of 55%, and those in the upper half were 75% correct.

To highlight consequences of this substantial variability in initial knowledge, we compared estimated opportunities needed to reach 80% mastery for students in the bottom and top halves of initial knowledge (see the second column of Table 2). We used the same formula for computing opportunities given above but

[‡] We report parameter estimate variability using interquartile range (the difference between the 75th and 25th percentiles) to reduce dependence on distributional assumptions inherent in other measures of variability.

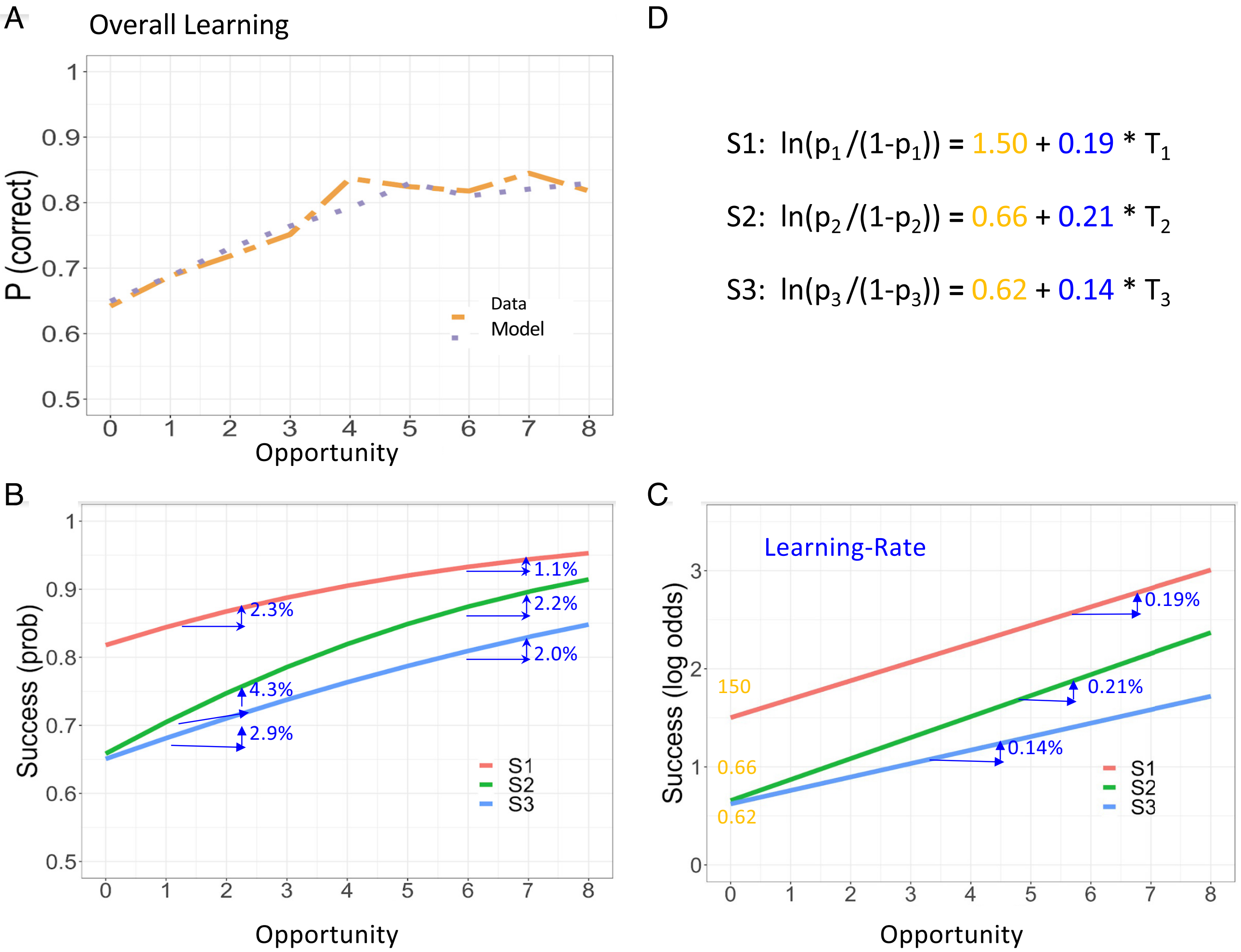

Fig. 2. Example learning curves from dataset 394. (*A*) Learning curve and model predictions average over all students and KCs. (*B*) Model-based learning curves for three randomly selected students showing nonlinear percentage point slopes at two different opportunities. (*C*) Same curves in log odds scale with intercept values (yellow) and linear slopes (blue). (*D*) Student i predicted success (p_i) at opportunity T as a function of intercept and slope.

replaced the overall initial knowledge (θ) with the 25th and 75th percentiles of the student initial knowledge estimates (θ_i). Whereas a student in the bottom half of initial knowledge needs about 13.13 opportunities to reach mastery, a student in the top half needs about 3.66 opportunities. In other words, a typical low initial knowledge student will take more than three times longer to reach mastery than a typical high initial knowledge student—a large difference for students who have met course prerequisites and been provided verbal instruction.

Students are Astonishingly Similar in Learning Rate. Whereas initial knowledge varies substantially across students, we found learning rate to be astonishingly similar across students. This contrast can be seen in model-based student learning curves, like the one shown above in Fig. 2*C*. The top of Fig. 3 shows such curves for four datasets representing different course content, educational levels, and kinds of educational technology. See *SI Appendix*, Figs. S6–S10 for KC and simulated data learning curves of all 27 datasets. Variation in initial knowledge is indicated by the wide range of intercepts in these curves. The similarity in student learning rate is illustrated by how generally parallel these curves are. While there are some cases of variation (e.g., see some nonparallel lines in the fourth panel for ds372), the log-odds increase in performance per opportunity is strikingly similar for most students in most datasets. This similarity in student learning rate is not only in contrast to much greater variation in student initial knowledge, but also in contrast to greater variation in knowledge component (KC) learning rates, shown in the middle of Fig. 3. This learning-rate variation by KC helps to alleviate a concern that we do not see variation in student learning rate because either our data or model are insufficient to detect such variation. The fact that we see substantial learning-rate variation by KC and the obvious variation in the simulated student curves (bottom row of Fig. 3) indicates learning-rate variations are detectable in these datasets with our model of learning, which relies on an empirically refined cognitive model of domain competence inserted into a mixed effects logistic regression growth model.

Low student learning-rate variation is common across all the datasets as indicated by the mostly parallel lines in Fig. 4, which shows the learning curves for the remaining 23 datasets. These curves also illustrate the much larger variation in student initial knowledge that is consistent across datasets. Note that the simulation results shown in Supplementary Information *SI Appendix*, Figs. S6–S10 (third column) confirm that our modeling approach can detect high student learning-rate variation (and low student initial knowledge variation) when it is present.

To gain a statistical sense for relative variation in learning-rate estimates, we calculated the SD and interquartile range of learning rate for students (δ_i) and KCs (γ_k) for each dataset. The median SD across datasets for student learning rate was 0.015 ($M = 0.022$;

Table 2. Median (SDs) across datasets for initial accuracy and opportunities to reach mastery for low (25) and high (75) percentiles of initial knowledge (assuming overall learning rate, *δ*) and learning rate (assuming overall initial knowledge, θ) for iAFM models

	Initial knowledge		Learning rate	
Percentile	Initial % correct	Opp to reach 80% mastery	Improvement % correct*	Opp to reach 80% mastery
25	55.21 (15.84)	13.13 (19.52)	1.70 (3.80)	7.89 (22.41)
50	66.05 (12.91)	6.54 (14.10)	2.25 (4.02)	7.27 (14.18)
75	75.17 (10.45)	3.66 (8.22)	2.56 (4.12)	6.94 (10.91)

*Difference between initial % correct and first opportunity: Improvements are linear in log odds but get smaller in percent increase as performance approaches 100% correct.

SD = 0.019) and the interquartile range was 0.018 (*M* = 0.028, *SD* = 0.024), whereas the median SD across datasets for KC learning rate was orders of magnitude larger at 0.102 (*M* = 0.175; *SD* = 0.177) and the interquartile range was seven times larger at 0.132 (*M* = 0.175, *SD* = 0.147). To help establish this low variability in student learning rate is robust and not, for example, only present when KC learning rate captures its variability, we investigated the impact of eliminating the KC learning-rate parameters. We found that the individual learning-rate estimates remain quite similar producing little change in their variability with an interquartile range of 0.021 instead of the 0.018 log odds for the full model.

Returning to Table 2**,** we provide a concrete sense of the small variability of student learning rate relative to variability in students' initial knowledge. For columns 3 and 4, we divided students into groups based on percentiles of student learning-rate estimates within each dataset (whereas columns 1 and 2 are divided based on percentiles of student initial knowledge estimates). In percentage terms, the interquartile range in variation for student learning rate (see column 3) is only about 1% per opportunity (2.56 to 1.70%), whereas the variation in initial knowledge is about 20% (75.17 to 55.21%). We calculated for each percentile of student learning rate how many opportunities a student needed to reach mastery by subtracting the overall initial knowledge (θ) for each dataset from the mastery criteria (80% = 1.4 log odds) and dividing it by the median student learning-rate parameter (δ_i) for that group of students (i.e., for each percentile of learning rate). Column 4 indicates that a typical student in the bottom half of learning rate (a slower learner) requires about 8 (Median = 7.89) opportunities to reach mastery, whereas a typical student in the top half of learning rate (a faster learner) requires about

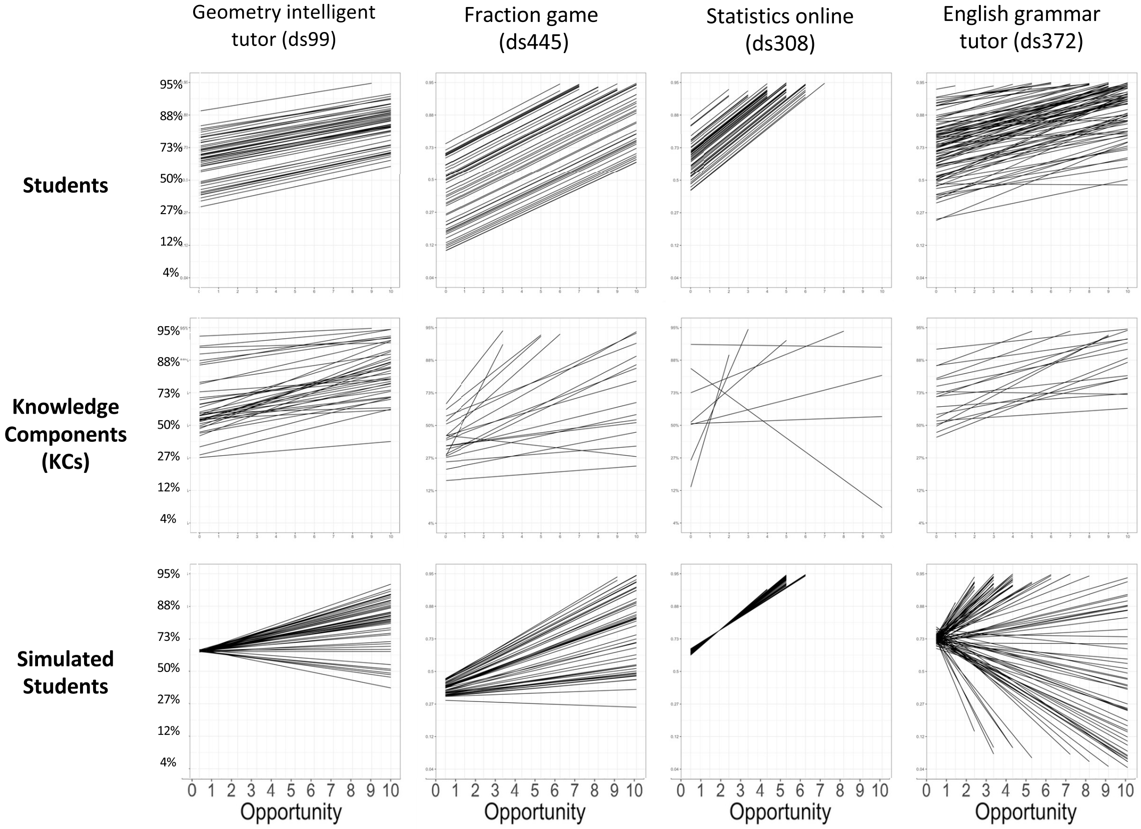

Fig. 3. Learning curves relating opportunities to practice to performance accuracy in percent correct displayed on a log odds scale. The top graphs show student learning curves indicating little variation in student learning rate (i.e., lines are mostly parallel) in contrast to large variation in initial performance. The middle graphs show knowledge component (KC) learning curves indicating that learning-rate variation is possible and measurable as these lines are not parallel. The bottom graphs demonstrate that the model can accurately identify high student learning-rate variation and low student initial performance variation when they are known to be present as determined by simulation.

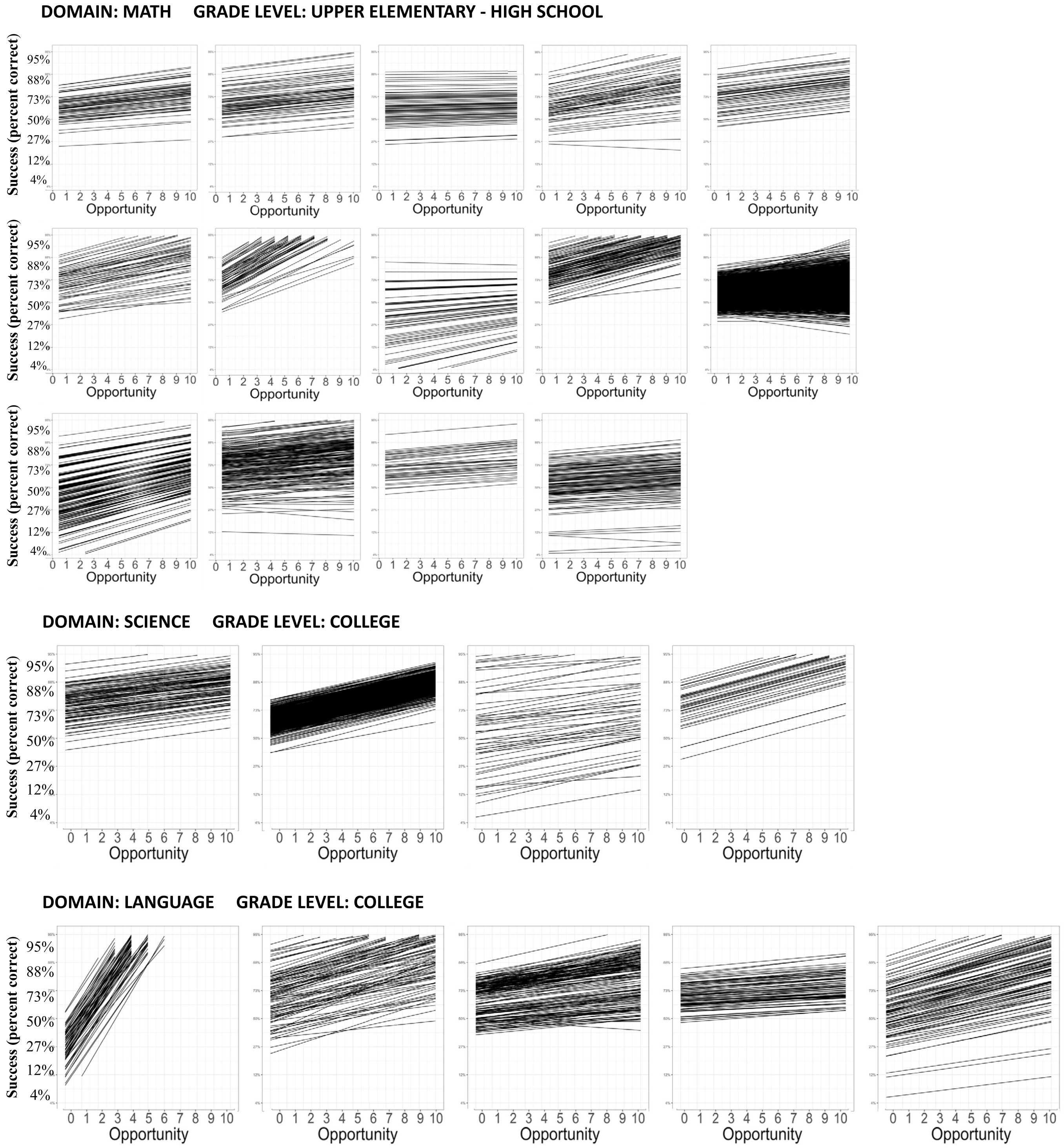

Fig. 4. Across three domains, and various grade levels, little variation is observed in student learning rate, but large differences are observed in student initial knowledge. These learning curves relate opportunities to practice to performance accuracy in percent correct displayed on a log odds scale.

7 (Median = 6.94) opportunities. In other words, a typical slower learner needs only one extra opportunity to keep pace with a typical faster learner. In contrast, we observed much larger differences in initial performance, with the bottom half of initial performance being about 10 opportunities behind the top half (13.13 to 3.66). The one opportunity difference to keep pace (i.e., span the interquartile range) in learning rate is an order of magnitude smaller than the 10-opportunity difference to catch up (i.e., span the interquartile range) in initial knowledge.

Perhaps there are particular circumstances in which higher student learning-rate variation is revealed. For example, perhaps high initial performers cannot demonstrate their learning potential on easier KCs, and higher learning rates would be revealed when we measure their rate on only the more difficult KCs. Conversely, low initial performers may show higher rates on easier KCs that are more within their reach. In fact, when we performed such an analysis, we did not see such a pattern. The estimated student learning rates remained quite similar whether measured using only hard KCs, only easy KCs, or all KCs. *SI Appendix* provides more detail.

To further explore the possibility of high student learning-rate variation in particular circumstances, we analyzed such variation disaggregated by domain (six language, 16 math, and five science) and by student subpopulation based on grade level (eight elementary, eight middle/high, 11 college). We use the interquartile range of the student learning-rate hyperparameter estimates within each dataset and compute medians for the domain and grade level subgroups. Interestingly, there appears to be greater student learning-rate variation in the six language domain datasets (median interquartile range of 0.045 log odds) than in the 16 math and five science domain datasets (median interquartile range of 0.017 log odds and 0.008 log odds, respectively). Student learning-rate variation due to grade levels is more consistent with median interquartile ranges of 0.015 log odds for upper elementary, 0.018 log odds for middle/high school, and 0.018 log odds for college.

Discussion

We set out to use amassed fine-grained, longitudinal data to better understand the progress of academic learning. While hoping to create a method to identify high-ability learners and understand their characteristics, we instead found strikingly similar rates of learning across students in the context of favorable conditions of interactive educational technologies. Along the way, we also demonstrated three other regularities. First, across a variety of courses, we found that initial practice performance is quite modest, about 65% correct (i.e., a failing grade), despite the general availability of up-front verbal instruction, such as lectures and readings. Second, we found that reaching a reasonable level of mastery (80% correct) requires substantial repeated practice, typically about seven practice opportunities. These results are consistent with learning theories suggesting induction from examples and doing is prominent in human learning (12, 41). Third, students' initial performance is highly variable despite students entering the courses in which the data were collected having met prerequisite requirements (for college courses) or age-level requirements (for K-12 courses) and having received verbal instruction.

Might the near constant student learning rate we observe be a consequence of limitations of our data or the measurement model? The first two results, that practice is needed and opportunity-based practice produces mastery, suggest our data have sufficient performance change to assess learning rate. The high variability in student initial performance indicates that our mixed effects growth model (iAFM) can differentiate individual student differences if present. The high variability in learning rate by knowledge component indicates that iAFM can differentiate learning-rate differences if present. Thus, limitations in the data or measurement model do not appear to be good explanations for our near constant student learning-rate observation.

Some readers may object that near constant student learning rate unrealistically implies that everyone can master advanced level calculus or interpret abstract data. Indeed, not everyone has favorable learning conditions nor will everyone choose to engage in the substantial number of practice opportunities required. However, our results suggest that if a learner has access to favorable learning conditions and engages in the many needed opportunities, they will master advanced level calculus. Other readers may object with intuitions of learning without substantial opportunities such as "I learned English without any practice" or "I learned calculus without attending lecture." These claims may overestimate achievement and underestimate implicit learning-by-doing (i.e., speaking English or solving calculus problems) and informal learning outside of school (e.g., a discussion of calculus with a friend).

While our investigation of 27 datasets covers a wide variety of academic settings, it is possible that our results do not generalize to all academic situations that provide favorable learning conditions. Research supports the favorability of repeated practice with quality feedback, as these educational technologies provide; however, there are forms of feedback such as peer interactions or classroom dialogue that are not well represented in our datasets.

Isolating Learning-Rate Measurement. Intuitions that different students learn at different rates may derive from contexts that do not provide the same level of controlled investigation under favorable learning conditions that the interactive learning technologies helped achieve. The striking regularity in learning rate revealed here is not apparent if we do not control or account for other factors that drive student performance. We must isolate performance changes due to particular opportunities for learning that are equivalent in nature and in number.

One problematic conceptualization is to measure learning rate in terms of calendar time, such as 10 years to become an expert (6). While convenient, this conceptualization ignores that some individuals will have many more learning opportunities than others in the same period of time. Using learning data captured through interactive learning technologies, we are able to get a more accurate accounting of learning opportunities than has been previously possible. And, indeed, we find these learning opportunities are much more predictive of learning outcomes than calendar time (a time-based model, time-AFM, systematically provides poor predictive fit).

A second problematic conceptualization of learning rate compares individuals getting different kinds of instructional support. To be sure, there is plenty of evidence that any of a wide variety of instructional interventions can greatly enhance student learning (e.g., refs. 25 and 42). However, such differences are a consequence of better instructional design not differences in students. Interactive learning technologies provide an effective way to control the nature of the learning opportunities that students experience. As driven by the materials and algorithms inherent in these technologies, all students are receiving the same essential instructional interactions.

A third problematic conceptualization is to merely compare individual performance overall without accounting for stable differences in performance due particularly to differences in prior experience. Our statistical modeling approach accounts for stable differences in performance by including an overall student performance latent variable as a per student baseline. We then measure learning rate as per-opportunity performance improvements above this baseline.

A fourth problematic conceptualization is to measure learning at the level of a broad topic or domain—as though learning opportunities exercise a general faculty (e.g., "math" or "scientific reasoning"). Instead, we model and measure learning in terms of finer grained components of knowledge. This knowledge component (KC) modeling has been demonstrated to more accurately predict human learning data than a general faculty approach (4). Thus, critical to our approach was to identify datasets with KC models of reasonable quality whereby researchers have used empirical methods to evaluate and refine these KC models. In fact, when we fit the statistical model (iAFM) to the same datasets using less accurate KC models, the estimates of overall learning rate go down and the estimates of student learning-rate variation are even smaller (*SI Appendix*). This analysis indicates the importance of KC model refinement (4, 35) in making student rate variation detectable. In fact, when we use a poor KC model, particularly the single KC model illustrated as Q0 in Table 1, the model with learning rate (iAFM) never outperforms the model without it (AFM).

Implications for Precise, Computational Theories of Learning. Taken together our findings pose a useful challenge for precise theories of human learning. Despite great progress in recent years, neither Cognitive Neuroscience nor Artificial Intelligence has provided a full, precise account of how humans learn complex academic concepts and skills. Learning curves have been used as an inspiration for learning theory development, but past work focused on modeling decrease in response times as a consequence of practice (43, 44). Such theory might be sufficient if performance accuracy was easy to achieve and most learning occurred as a speed-up in accurate performance. Our results suggest otherwise. Given our findings that a typical student starts practice-based learning at about 65% accuracy and that substantial practice, typically about seven opportunities per knowledge component, are needed to achieve 80% accuracy, it is clear that learning theory must also account for changes in performance accuracy.

Our results suggest three specific theoretical challenges. A precise learning theory should explain and account for 1) continued changes in performance accuracy due to deliberate practice opportunities after initial up-front verbal instruction, 2) substantial student variation in initial performance, and 3) much smaller variation in student learning rate across practice opportunities. We briefly discuss each.

Explaining deliberate practice benefits. That up-front lectures and readings seem to produce limited performance accuracy is surprising given the great efforts educators continue to put into producing lectures and texts and given that most learners advocate explicit learning as the best way to learn (45). Books and then recorded lectures have facilitated broader dissemination of knowledge historically, but much emphasis on lecture recording remains today even in online course contexts where interactive practice is feasible and effective (cf., 20). A theoretical postulate consistent with limited accuracy after up-front verbal instruction is that human learning is not simply about the explicit processing, encoding, and retrieval of verbal instruction but as much or more about implicit or nonverbal learning-by-doing in varied practice tasks where interactive feedback is available (12).

We observed that giving learners well-designed practice opportunities with feedback produces performance accuracy increases but learners typically require many such opportunities. Such interactive practice-based theory has support from empirical studies of expertise development (5), experimental studies of the testing effect (17), and active learning (46). Theoretical models of human skill acquisition are generally consistent in the qualitative prediction that many learning opportunities are needed for a skill to be accurately acquired or a fact to be robustly recalled at a long interval (47, 48).

Explaining big differences in student initial knowledge intercepts. A class of general computational theories of cognitive skill acquisition (e.g., 47, 49, 50, 51, 52) suggest that expertise develops through experience. This experience produces new skills stored in a procedural memory system made up of conditional knowledge components, typically implemented as if-then production rules. Such theories provide a straightforward explanation of big differences in student prior domain knowledge, namely, that some students have had more domain-relevant prior experiences than others. These past experiences produce domain-specific learning opportunities before the window of observation within our datasets (e.g., discussion of fractions with a parent while making pancakes at home before fractions are covered in school).

This prior-opportunities hypothesis has been used to explain large intercept differences in reaction time learning curve data between children (with higher initial reaction times) and adults (with lower initial reaction times) given repeated practice on mental rotation tasks (53). The learning curves of the two groups match under the single assumption that adults have had about 2,000 more prior opportunities to practice than the children. The Apprentice Learner (AL; ref. 54) theory has been used to support this prior-opportunities hypothesis in making accurate predictions of individual intercept differences in error-rate learning curves. As a fully functional computational model of learning, AL learns from tutoring interactions like those provided by the interactive educational technologies used to produce our datasets and thus makes these predictions with only a single parameter per student representing unobserved prior practice. Unlike statistical models (e.g., refs. 43, 55, and 56), AL does not need parameters for knowledge component difficulty and learning rate because such differences are emergent from declining performance errors produced by the learning mechanisms inherent in it (e.g., learning how to do things to produce the then-part of production rules and learning both where to get needed information and when to do things to produce the if-part of production rules).

The prior-opportunities hypothesis suggests a concrete, though challenging, empirical test: If researchers can track and count domain-relevant learning opportunities that students experience prior to course entry, they should find a) large differences across students in these prior opportunities and b) that these differences substantially account for large initial performance differences at the start of within-course practice that we have observed.

Toward explaining small differences in student learning rate. We can infer a prediction of student learning-rate variation from the learning mechanisms posited in prior skill acquisition theories (e.g., 47, 49). The compilation learning mechanism (49) posits that domain-specific skills are acquired from preexisting domain-specific declarative knowledge and from preexisting domain-general procedural knowledge that interprets these declarative memories to form new domain-specific skills. If we assume individual variation in the quality or quantity of this domain-general background knowledge, it follows from these skill acquisition theories that we should observe individual differences in learning rate. Similar arguments follow from the chunking mechanism (57) and inductive learning mechanisms in the Apprentice Learner (53).

Such background knowledge is used indirectly in the domain as support to answer questions (e.g., negative number concepts in equation solving) or as part of processing learning materials (e.g., reading skills for comprehending solution directions in equation solving). It can be distinguished from prior domain knowledge, which is used directly to answer questions or perform reasoning steps in the instructional domain (e.g., adding to both sides in algebra equation solving). Large differences in prior opportunities, it would seem, should not only produce differences in prior domain knowledge, as discussed above, but also differences in background knowledge. In turn, differences in background knowledge should produce differences in learning rate (cf., 58, 59).

While we did find large differences in initial knowledge, we found quite small differences in learning rate. Why might that be? We suggest a disjunctive learning path hypothesis based on our observations of learning processes of the Apprentice Learner (AL) and consistent with skill acquisition mechanisms in other theories (47, 49). AL specifies a mechanism of academic learning whereby learners use background knowledge to search for and induce mental derivations or explanations of examples they experience (cf., 60, 61). To be sure, these mental explanations are modeled and conceived of as mostly nonverbal, inductive brain processes (cf., 12) not explicit verbal reasoning that the term "explanation" may evoke. AL produces many alternative explanations of the same example steps especially when simulated learners are given different subsets of that background knowledge. Thus, AL predicts that, with different sampling of background knowledge, student learning produces differences in mental representation but similar performance outcomes. For example, one student may have existing background knowledge to self-explain an algebra example involving negative number subtraction, whereas another student does not but compensates with a conceptual strategy using a number line to self-explain the same example.

Consistent with this disjunctive learning path hypothesis, a neuroimaging study demonstrated that students achieved equivalent performance in math problem solving with quite different mental representations (62). Students instructed with a verbal representation solved problems as effectively but with higher activation in the left-prefrontal cortex than students instructed through a symbolic representation, who revealed higher activation in the bilateral parietal cortices.

The higher learning-rate variation we observed in the language datasets than in the math and science datasets also appears consistent with the disjunctive learning path hypothesis. The math and science domains allow for multiple learning paths in that the subject-matter includes generalized skills and rediscoverable principles and fewer verbatim facts than the language domains (cf., 12). Learning in language domains is thus more dependent on rote memory to acquire arbitrary mappings (e.g., in English, oceans are referenced using "the" whereas lakes are not). Thus, variations in rote memory processing may produce greater learning-rate variation in those domains than in math and science domains where rote memory limitations can be supplemented or compensated with general skill induction or sense making processes.

More generally, our educational system may be reasonably uniform in providing students with sufficient background knowledge for learning such that, for example, students enter an algebra course with enough background knowledge of integers, rationals, and arithmetic to learn from good examples, practice, and instructional feedback. Moreover, given favorable learning conditions, student learning may be substantially robust to small gaps in background knowledge. With good instruction, such as the quality deliberate practice that interactive learning technology systems provide, students can compensate during learning for some gaps. We suggest further theory development and learning curve modeling to test these hypotheses.

Practical Implications. The learning-rate question is practically important because it bears on fundamental questions regarding education and equity. Can anyone learn to be good at anything they want? Or is talent, like having a "knack for math" or a "gift for language," required? Our evidence suggests that given favorable learning conditions for deliberate practice and given the learner invests effort in sufficient learning opportunities, indeed, anyone can learn anything they want. If true, this implication is good news for educational equity—as long as our educational systems can provide the needed favorable conditions and can motivate students to engage in them. The variety of well-designed interactive online practice technologies used to produce our datasets point to a scalable strategy to provide these favorable conditions. Importantly, these technologies were well engineered to provide the key features of deliberate practice including well-tailored task design, sufficient repetition in varied contexts, feedback on learners' responses, and embedded instruction when learners need it. At the same time, students do not learn from these technologies if they do not use them. Recent research providing human tutoring to increase student motivation to engage in difficult deliberate practice opportunities suggests promise in reducing achievement gaps by reducing opportunity gaps (63, 64).

Materials and Methods

Datasets. This project was possible because of LearnLab's DataShop, the world's largest repository of student learning data (65). We used 27 datasets (Table 2 see 66) from DatatShop that include an assortment of domains (e.g., geometry, fractions, physics, statistics, English articles, Chinese vocabulary), of educational levels (e.g., grades 5 to 12, college, adult learners), and of settings (e.g., in class vs. out of class as homework). In general, students worked at their own pace through course materials and received as-needed assistance in the form of hints and feedback. In many cases, a predetermined period of time was set for completing the work (e.g., one or more class periods).

Within these educational technologies, students perform tasks by answering questions or, in some cases, entering solutions to problems in a step-by-step fashion. All entries are either selected responses or short constructed responses that are automatically scored, sometimes with the help of Artificial Intelligence algorithms. Whereas some student responses are to four-choice multiple choice questions, most requested student responses are open. Such responses include text fields where students enter numbers (e.g., "72.3"), expressions (e.g., "(972+b)/5"), or a word (e.g., the pinyin spelling of a Chinese symbol). They also include graphical user interface actions such as clicking a place on a number line (ds445) or drawing a force vector (ds104). Some multiple choice questions involve more than four options such as a list of some 12 possible explanations for an English article choice. Example tasks can be found in *SI Appendix* (*SI Appendix*, Figs. S1–S5). Student responses to these tasks are automatically tagged as correct when students answer correctly on their first attempt without asking for a hint. Otherwise, the task response is tagged as an error. To estimate performance at a given task opportunity, only the student's first attempt is considered, though subsequent student attempts and system feedback are critical contributors to learning. We define learning as a positive change in performance and operationalize learning as a reduction in error rate (or increase in correctness rate) over successive opportunities to perform a task associated with a specific knowledge component.

Dataset Selection Criteria. Among hundreds of datasets available in DataShop, we identified 27 datasets to include in our analysis (*SI Appendix*, Tables S1 and S2). To achieve accurate parameter estimation from a dataset, it is critical to have a quality knowledge component (KC) model (cf., 1, 4). Thus, we looked for datasets where a KC model meets a set of precise criteria for quality and interpretability. To be included, a dataset must have an associated KC model that is better than at least two extreme alternatives (defined as lowest root mean squared error on the test set in threefold item-blocked cross validation), an item-based model where each distinct task is coded as a different KC (e.g., Q3 in Table 1), and a faculty model where all unit tasks are coded as a single KC (e.g., Q0 in Table 1). For straightforward interpretation, we only considered KC models that involve a single KC label per unit task (see footnote 1). For datasets that had more than the three KC models implied above, we selected the KC model that had the best overall prediction fit in item stratified cross validation (always comparing on the same sample of student observations).

We investigated the impact that the quality of the KC model had on the results we present. Better KC models tend to both increase overall learning rate (cf., 12) and slightly increase student learning-rate variability. Nevertheless, our main results, particularly low student learning-rate variability, remain regardless of the KC model chosen (see *SI Appendix* for details).

We eliminated two datasets where the initial overall success rate was greater than 80% as these datasets leave less room to observe learning. Among the selected 27 datasets, we noticed that some specific KCs were too limited in number of associated data points. Thus, in each dataset, a KC was included only if there were data from at least 10 students with at least two learning opportunities. For 13 datasets, all KCs were included and in the other 14 datasets, an average of 2.7 of 952 KCs were excluded.

Data, Materials, and Software Availability. All data are available from links in *SI Appendix*, Table S3 and all analysis scripts are available here: https://pslcdatashop.web.cmu.edu/Files?datasetId=4629 (66).

ACKNOWLEDGMENTS. We thank Carnegie Learning, Inc., for making available a number of the datasets used in this study and the National Science Foundation for support in creating the data infrastructure and performing the research including grants from Social, Behavioral and Economic Sciences (SBE 0354420 and SBE 0836012), from Advanced Cyberinfrastructure (OAC 1443068) and from Behavioral and Cognitive Sciences (BCS 1824257).

1. H. Cen, K. R. Koedinger, B. Junker "Learning Factors Analysis: A general method for cognitive model evaluation and improvement" in *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, M. Ikeda, K. D. Ashley, T.-W. Chan, Eds. (Springer-Verlag, Berlin, 2006), pp. 164–175.
2. R. Liu, K. R. Koedinger, Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains. *J. Educ. Data Mining* **9**, 25–41 (2017).
3. K. R. Koedinger, E. A. McLaughlin "Seeing language learning inside the math: Cognitive analysis yields transfer" in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, S. Ohlsson, R. Catrambone, Eds. (Austin, TX, 2010), pp. 471–476.
4. K. R. Koedinger, M. V. Yudelson, P. I. Pavlik, Testing theories of transfer using error rate learning curves. *Topics Cogn. Sci.* **8**, 589–609 (2016).
5. K. A. Ericsson, R. T. Krampe, C. Tesch-Romer, The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* **100**, 363–406 (1993).
6. K. A. Ericsson, Deliberate practice and acquisition of expert performance: a general overview. *Acad. Emerg. Med.* **15**, 988–94 (2008).
7. B. N. Macnamara, M. Maitra, The role of deliberate practice in expert performance: revisiting Ericsson, Krampe & Tesch-Römer (1993). *Roy. Soc. Open Sci.* **6**, 190327 (2019).
8. B. N. Macnamara, D. Moreau, D. Z. Hambrick, The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspect. Psychol. Sci.* **11**, 333–350 (2016).
9. National Research Council, "Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools" (The National Academies Press, Washington, DC, 2002).
10. C. L. Zerr *et al.*, Learning efficiency: Identifying individual differences in learning rate and retention in healthy adults. *Psychol. Sci.* **29**, 1436–1450 (2018).
11. B. S. Bloom, *Human Characteristics and School Learning* (McGraw-Hill, 1976).
12. K. R. Koedinger, A. C. Corbett, C. Perfetti, The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cogn. Sci.* **36**, 757–798 (2012).
13. K. R. Koedinger, C. Perfetti, Pittsburgh Science of Learning Center Strategic Plan. Available at https://pslcdatashop.web.cmu.edu/Files?datasetId=4629 (2011).
14. R. Liu, K. R. Koedinger "Towards reliable and valid measurement of individualized student parameters" in *Proceedings of the 10th International Conference on Educational Data Mining*, X. Hu, T. Barnes, A. Hershkovitz, L. Paquette, Eds. (Wuhan, China, 2017), pp. 135–142.
15. B. S. Bloom, *Learning for Mastery* (University of California Press, Los Angeles, USA, 1968).
16. N. Yannier, S. E. Hudson, K. R. Koedinger, AI from the screen into the Physical World. *Science* **374**, 26–27 (2021).
17. H. L. Roediger III, J. D. Karpicke, Test-enhanced learning: Taking memory tests improves long-term retention. *Psychol. Sci.* **17**, 249–255 (2006).
18. P. Black, W. Dylan, In praise of educational research: Formative assessment. *Br. Educ. Res. J.* **29**, 623–637 (2003).
19. B. N. Macnamara, D. Z. Hambrick, F. L. Oswald, Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychol. Sci.* **25**, 1608–1618 (2014).
20. K. R. Koedinger, J. Kim, J. Jia, E. A. McLaughlin, N. L. Bier "Learning is not a spectator sport: Doing is better than watching for learning from a MOOC" in *Proceedings of the Second ACM Conference on Learning at Scale* (2015), pp. 111–120.
21. J. Hattie, H. Timperley, The power of feedback. *Rev. Educ. Res.* **77**, 81–112 (2007).
22. B. Wisniewski, K. Zierer, J. Hattie, The power of feedback revisited: a meta-analysis of educational feedback research. *Front. Psychol.* **10**, 1–14 (2020).
23. R. Moreno, Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instr. Sci.* **32**, 99–113 (2004).
24. J. K. Crissman, *The Design and Utilization of Effective Worked Examples: A Meta-analysis* (ETD collection for University of Nebraska, Lincoln, AAI3208114 (2006).
25. H. Pashler "Organizing Instruction and Study to Improve Student Learning IES Practice Guide" (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.
26. R. J. C. M. Salden, K. R. Koedinger, A. Renkl, V. Aleven, B. M. McLaren, Accounting for beneficial effects of worked examples in tutored problem solving. *Educ. Psychol. Rev.* **22**, 379–392 (2010), 10.1007/s10648-010-9143-6.
27. K. R. Koedinger, J. C. Stamper, E. A. McLaughlin, T. Nixon "Using data-driven discovery of better student models to improve student learning" in *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, H. C. Lane, K. Yacef, J. Mostow, P. Pavlik, Eds. (Memphis, TN, 2013), pp. 421–430.
28. R. Liu, K. R. Koedinger, E. A. McLaughlin "Interpreting model discovery and testing generalization to a new dataset" in *Proceedings of the 7th International Conference on Educational Data Mining*, J. Stamper, Z. Pardos, M. Mavrikis, B. M. McLaren, Eds. (London, UK, 2013), pp.107–113.
29. K. A. Ericsson, Ed., *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games* (Psychology Press, ed. 1, 1996).
30. F. G. W. C. Paas, J. J. G. van Merriënboer, Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *J. Educ. Psychol.* **86**, 122–133 (1994).
31. J. R. Busemeyer, A. Diederich, *Cognitive Modeling* (Sage Publishing, 2009).
32. T. Barnes "The Q-matrix method: Mining student response data for knowledge" in *Proceedings of AAAI 2005 Educational Data Mining Workshop* (2005).
33. M. C. Desmarais, R. Naceur, R. "A Matrix Factorization Method for Mapping Items to Skills and for Enhancing Expert-Based Q-matrices" in *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, H. C. Lane, K. Yacef, J. Mostow, P. Pavlik, Eds. (Memphis, TN, 2013), pp. 441–450.
34. J. C. Stamper, K. R. Koedinger "Human-machine student model discovery and improvement using data" in *Proceedings of the 15th International Conference on Artificial Intelligence in Education*, G. Biswas, S. Bull, J. Kay, A. Mitrovic, Eds. (Auckland, New Zealand, 2011), pp. 353–360.
35. K. R. Koedinger, E. A. McLaughlin, J. C. Stamper "Automated student model improvement" in *Proceedings of the 5th International Conference on Educational Data Mining*, K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, J. Stamper, Eds. (Chania, Greece, 2012), pp. 17–24.
36. M. Wilson, P. de Boeck, "Descriptive and explanatory item response models" in *Explanatory Item Response Models*, P. de Boeck, M. Wilson, Eds. (Springer, 2004), pp. 43–74.
37. Y. Bergner "Model-based collaborative filtering analysis of student response data: Machine-learning item response theory" in *Proceedings of the 5th International Conference on Educational Data Mining*, K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson, J. Stamper, Eds. (Chania, Greece, 2012), pp. 95–102.
38. P. Pirolli, M. Wilson, A theory of the measurement of knowledge content, access, and learning. *Psychol. Rev.* **105**, 58–82 (1998).
39. H. Cen, K. Koedinger, B. Junker "Comparing two IRT models for conjunctive skills" in *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS 2008) Lecture Notes in Computer Science*, B. P. Woolf, E. Aimeur, R. Nkambou, S. Lajoie, Eds. (Montreal, Canada, 2008), pp. 796–798.
40. E. J. Wagenmakers, A practical solution to the pervasive problems of*p* values. *Psychon Bull Rev.* **14**, 779–804 (2007).
41. J. H. Holland, K. J. Holyoak, R. E. Nisbett, P. Thagard, *Induction: Processes of Inference, Learning, and Discovery* (The MIT Press, 1989).
42. K. R. Koedinger, J. L. Booth, D. Klahr, Instructional complexity and the science to constrain it. *Science* **342**, 935–937 (2013).
43. A. Heathcote, S. Brown, D. J. K. Mewhort, The power law repealed: The case for an exponential law of practice. *Psychonomic Bull. Rev.* **7**, 185–207 (2000).
44. A. Newell, P. S. Rosenbloom, "Mechanisms of skill acquisition and the power law of practice" in *Cognitive Skills and Their Acquisition*, J. R. Anderson, Ed., (Erlbaum, 1981), pp. 1–55.
45. P. F. Carvalho, E. A. McLaughlin, K. R. Koedinger "Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes" in *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (2017), pp. 204–209.
46. L. Deslauriers, L. S. McCarty, K. Miller, K. Callaghan, G. Kestin "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom" *in Proceedings of the National Academy of Sciences* (2019), pp. 19251–19257.
47. J. E. Laird, *The Soar Cognitive Architecture* (MIT Press, 2012).
48. J. R. Anderson, *How Can the Human Mind Occur in the Physical Universe?* (Oxford University Press, 2009).
49. J. R. Anderson, C. Lebiere, *The Atomic Components of Thought* (Lawrence Erlbaum Associates, Mahwah, NJ, 1998).
50. P. Langley, D. Choi "A unified cognitive architecture for physical agents" in *Proceedings of the Twenty-first AAAI conference on Artificial Intelligence* (AAAI Press, Boston, MA, 2006), pp. 1469–1474.
51. K. VanLehn "Human procedural skill acquisition: Theory, model, and psychological validation" in *Proceedings of the Third National Conference on Artificial Intelligence* (Washington, D.C., 1983), pp. 420–423.
52. C. J. MacLellan, E. Harpstead, R. Patel, K. R. Koedinger "The Apprentice Learner architecture: Closing the loop between learning theory and educational data" in *Proceedings of the 9th International Conference on Educational Data Mining*, T. Barnes, M. Chi, M. Feng, Eds. (2016), pp. 151–158.
53. R. Kail, Y. S. Park, Impact of practice on speed of mental rotation. *J. Exp. Child. Psychol.* **49**, 227–244 (1990).
54. D. Weitekamp III, Z. Ye, N. Rachatasumrit, E. Harpstead, K. R. Koedinger, "Investigating differential error types between human and simulated learners" in *Lecture Notes in Computer Science AIED 2020*, I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, E. Millán, Eds. (Springer, Cham, 2020), **vol. 12163**, pp. 586–597.
55. N. Li, E. Stampfer, W. Cohen, K. Koedinger "General and efficient cognitive model discovery using a simulated student" in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Berlin, Germany, 2013), vol. **35**, pp. 894–899.
56. M. Steyvers, G. E. Hawkins, F. Karayanidis, S. D. Brown, A large-scale analysis of task switching practice effects across the lifespan. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 17735–17740 (2019).
57. J. E. Laird, P. S. Rosenbloom, A. Newell, Chunking in Soar: The anatomy of a general learning mechanism. *Mach. Learn.* **1**, 11–46 (1986).
58. A. Delahay, M. Lovett "Distinguishing two types of prior knowledge that support novice learners" in *Proceedings of the 41st Annual Conference of the Cognitive Science Society 2019*, A. K. Goel, C. M. Seifert, C. Freksa, Eds. (Montreal, QC, 2019), pp. 1620–1626.
59. N. Li, N. Matsuda, W. Cohen, K. Koedinger "Towards a computational model of why some students learn faster than others" in *Proceedings of the AAAI 2010 Fall Symposium on the Cognitive and Metacognitive Educational Systems* (Arlington, VA, 2010), pp. 40–46.
60. R. K. Atkinson, A. Renkl, M. M. Merrill, Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *J. Educ. Psychol.* **95**, 774–783 (2003).
61. T. M. Mitchell, R. M. Keller, S. T. Kedar-Cabelli, Explanation-based generalization: A unifying view. *Mach. Learn.* **1**, 47–80 (1986).
62. M. H. Sohn *et al.*, Behavioral equivalence, but not neural equivalence—neural evidence of alternative strategies in mathematical thinking. *Nat. Neurosci.* **7**, 1193–1194 (2004).
63. J. Guryan *et al.*, *Not Too Late: Improving Academic Outcomes among Adolescents* (National Bureau Econ. Res., Working Paper 28531 2021, March).
64. D. R. Chine, *et al.* "Educational equity through combined human-ai personalization: A propensity matching evaluation" in *Artificial Intelligence in Education, AIED 2022, Lecture Notes in Computer Science*, M. M. Rodrigo, N. Matsuda, A. I. Cristea, V. Dimitrova, Eds. (2022), pp. 366–377.
65. K. R. Koedinger *et al.*, "A data repository for the EDM community: The PSLC DataShop" in *Handbook of Educational Data Mining*, C. Romero, S. Ventura, M. Pechenizkiy, R. S. J. D. Baker, Eds. (CRC Press, 2010).
66. K. R. Koedinger, P. F. Carvalho, L. Ran, McLaughlin, Human Learning Rate. DataShop. https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=4629 (2020).