

**Mathematical learning models that depend on prior knowledge and instructional strategies**

David E. Pritchard and Young-Jin Lee

*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

Lei Bao

*Department of Physics, Ohio State University, Columbus, Ohio 43210, USA*

(Received 27 March 2007; published 20 May 2008)

We present mathematical learning models—predictions of student’s knowledge vs amount of instruction—that are based on assumptions motivated by various theories of learning: *tabula rasa*, constructivist, and tutoring. These models predict the improvement (on the post-test) as a function of the pretest score due to intervening instruction and also depend on the type of instruction. We introduce a connectedness model whose connectedness parameter measures the degree to which the rate of learning is proportional to prior knowledge. Over a wide range of pretest scores on standard tests of introductory physics concepts, it fits high-quality data nearly within error. We suggest that data from MIT have low connectedness (indicating memory-based learning) because the test used the same context and representation as the instruction and that more connected data from the University of Minnesota resulted from instruction in a different representation from the test.

DOI: [10.1103/PhysRevSTPER.4.010109](https://doi.org/10.1103/PhysRevSTPER.4.010109)

PACS number(s): 01.40.Fk, 01.40.Ha

**I. INTRODUCTION**

The stimulus for this paper was trying to understand what learning mechanism(s) could lead to Hake’s finding<sup>1</sup> that for prescores and postscore on “concept tests”<sup>2</sup> in Newtonian mechanics, 62 different classes tended to have similar average normalized gain that strongly depended on the type of instruction (traditional vs interactive) but insignificantly (correlation +0.02) on the prior knowledge of the class, as measured by the preinstruction score. In particular, why should the score increase from pretest to post-test be proportional to what the students initially do not know, rather than being the same for all students, or even increasing with their prior knowledge?

The central focus of this paper is to construct mathematical learning models that yield expressions for student knowledge vs instructional time based on clear assumptions that are motivated by various learning theories, then to find which model best fits observed student learning data—in particular, the variation of learning with the amount of students’ prior knowledge. The aim of this paper is thereby to provide a quantitative tool that allows the parametrization of measured learning data (e.g., from pretesting and post-testing before and after instruction) in terms of a new parameter (in addition to the normalized gain) that connects with learning theory discussions. We hope that this work will provide a way to determine what type of learning is actually occurring, thereby allowing both “learning type” and “amount learned” to be inferred from assessments made before and after a period of instruction.

Our key insight is that existing learning theories suggest that the rate of learning should have significantly different dependences on the student’s *prior knowledge*—what he does or does not know when he encounters each new thing to be learned. The first such theory is *tabula rasa*, the Aristotelian idea popularized by Locke that the mind can be seen as a blank slate which is imprinted with knowledge initially through experience (and later through reasoning as well).<sup>3</sup> The second, *constructivism*, is the notion that new knowl-

edge is “constructed” from associations involving prior knowledge<sup>4-7</sup> and thus that increased prior knowledge should positively affect the rate of learning. Finally, *tutoring* is one-on-one expert mentoring that is tailored to the particular student, e.g., as practiced by Socrates.<sup>8,9</sup>

These assumptions lead to quite different mathematical expressions for a student’s rate of learning. We solve the resulting differential equations to yield predictions for the student’s knowledge after a given amount of instruction. For several large college classes, we then show that based on preinstruction test scores (“pretest”), a hybrid learning model with a new “connectivity parameter” fits postinstruction test scores<sup>10</sup> nearly within error, while other models are excluded with certainty.

The resulting inferences about the connectedness of the learning to prior knowledge are in accord with expectations from studies of knowledge transfer:<sup>11</sup> memory-based learning (i.e., transfer from instruction to assessment) is inferred when instruction is based on questions similar to those on the test, whereas more connected learning is inferred when instruction uses a different representation than the test.

Our models are applicable to already accumulated pretest and post-test data, to year to year test results that are currently being accumulated in many school systems, and as a byproduct of data collected in online learning systems. They therefore offer a new way to make inferences about the *type* of learning that is occurring.

**II. MODELS**

The dependent variable in our models is the student’s knowledge after a specified amount of instruction. We divide the test domain,  $T$ , into a fraction  $K_T(t)$  that is known (i.e., tests “correct”), and what is unknown,  $U_T(t)$ . Here,  $t$  is the amount of instruction (i.e., teaching) the student has had in the test domain. The variable  $t$  is the amount of teaching or instruction that has occurred normalized by the total knowledge in the test domain and can exceed unity (e.g., the

teacher can review a material so that a particular knowledge nugget is taught more than once). There is also a “sticking coefficient”  $\alpha$ , which is the probability that something taught sticks in the student’s mind when all other conditions for potential learning within that particular model are present. In practice, only the product  $\alpha t$  appears in the models; we vary  $\alpha t$  to fit pre-post-testing data.

Each of our models starts with assumptions that yield a particular differential equation for  $dU_T(t)/dt$ , the rate of change of unknown knowledge in the test domain.  $dU_T(t)/dt$  is the negative of the learning rate since  $U_T(t) = 1 - K_T(t)$ . The differential equations for  $U_T(t)$  are simpler than those for  $K_T(t)$ , in part because instruction is generally considered as teaching students what they do not already know. For each learning model, we give an analytic solution for  $K_T(t)$  [and therefore for  $U_T(t)$ ].

Typically, our models will be applied to data consisting of prescores and postscores on concept tests for a class of students who undergo the same amount of instruction,  $t_{instr}$ . Each student has a different initial knowledge (pretest score)  $K_{T0} = K_T(t=0)$  (which could be attributed to prior instruction at negative  $t$ ). Each different learning model then gives  $K_T(t_{instr})$ , a prediction of the student’s post-test score. Such preinstruction and postinstruction scores are generally (at least in physics education research) summarized in terms of the “normalized gain,”<sup>1,12–14</sup> which the model predicts to be a function of the pretest score  $K_{T0} = K_T(t=0)$ ,

$$g(K_{T0}) = [K_T(t_{instr}) - K_{T0}] / [1 - K_{T0}]. \quad (1)$$

The normalized gain is the fraction of the unknown (i.e., wrong) knowledge on the pretest  $(1 - K_{T0})$  that is learned (i.e., correct) on the post-test after an amount of instruction  $t_{instr}$ .<sup>15–20</sup>

### A. Pure memory model

Our “pure memory model” is motivated by the tabula rasa theory of learning. The model assumes that the student’s memory (the slate) is blank in the unknown fraction of the test domain,  $U_T(t) = 1 - K_T(t)$ . The assumptions of the pure memory model are the following:

- (1) The instructional process sends out knowledge nuggets uniformly distributed in all respects over the test domain.
- (2) Obviously, only the fractions  $U_T(t)$  of these are unknown and hence can be learned.
- (3) The sticking probability of the nuggets is  $\alpha$  for all students.
- (4)  $\alpha$  does not have a systematic dependence on  $K$  or  $U$ .

This model applies to rote memorization since assumption (3) means that the probability of remembering the presented nuggets is independent of the student’s other knowledge and recent learning history.

With these assumptions,

$$dU_T(t)/dt = -\alpha_{memory} U_T(t), \quad (2)$$

which has the solution

$$U_T(t) = U_T(0) \exp[-\alpha_{memory} t],$$

leading to

$$K_T(t) = 1 - (1 - K_{T0}) \exp[-\alpha_{memory} t], \quad (3)$$

$$g(K_{T0}) = 1 - \exp[-\alpha_{memory} t_{instr}]. \quad (4)$$

The most important prediction of the pure memory model is that  $g$  is independent of the pretest score,  $K_{T0}$ . Thus, this model fits Hake’s data for different classes undergoing the same type of pedagogy.

We will subsequently refer to this model as the pure memory model to emphasize that it is based on the mathematical assumption of independence of learning on prior or current knowledge. The pure memory model has the same mathematical form as some Pavlovian learning models, such as stimulus sampling theory<sup>21</sup> and the Rescorla–Wagner model,<sup>22</sup> which model the change in behavior of subjects across repeated trials (instruction) at a task. Similar mathematics occurs in some stochastic models for learning.<sup>23</sup> For example, Eq. (3) is similar to Hull’s “habit strength,”<sup>24</sup> where  $\alpha_{memory}$  is the “learning rate parameter” and  $t$  is the “number of trials.” The assumptions behind all these models are much sharper than the many ideas associated with tabula rasa and generally boil down to the idea that subsequent instruction always has the same fractional learning effect (called “stationarity” in psychology and “current knowledge state” in our model).

### B. Simple connected model

Our simple connected model is motivated by the constructivist view that students learn new knowledge by constructing an association between it and some prior knowledge. This implies that the more prior knowledge one has, the faster learning will occur and, conversely, that if the relevant knowledge needed for the construction is unknown, then learning will be slowed. The learning rate is now proportional to three factors: the probability  $U_T(t)$  that the knowledge nugget strikes an unknown region, the probability  $K_T(t)$  that the appropriate connecting knowledge is already known, and the probability that the nugget will stick (i.e., the association will be constructed),  $\alpha_{connected}$ . The extra factor  $K_T(t)$  in the simple connected model reflects that the learner must possess the existing knowledge with which to connect an unknown knowledge nugget in order to learn it. [Our simplifying assumption is that one and only one such connection to knowledge within the test domain is needed; otherwise,  $K_T(t)$  would be raised to some power.] The model also makes the four assumptions of the pure memory model above, giving

$$dU_T/dt = -\alpha_{connected} U_T(t) K_T(t) = -\alpha_{connected} U_T(t) [1 - U_T(t)], \quad (5)$$

which has the solution

$$K_T(t) = \frac{1}{1 + (1 - K_{T0}) \exp[-\alpha_{connected} t] / K_{T0}} \quad (6)$$

(a logistic function),

$$g(K_{T0}) = \frac{K_{T0}(\exp[\alpha_{connected} t_{instr}] - 1)}{(1 - K_{T0}) + K_{T0} \exp[\alpha_{connected} t_{instr}]} \quad (7)$$

Interestingly, this model would apply to classes that receive instruction solely by peer to peer interactions involving random pairs of students if one assumes that learning occurs if and only if one student knows the answer to the given question and the paired student does not (and there is a uniform probability  $> \frac{1}{2}$  that the “right” rather than the “wrong” student’s view will be adopted by both). Under these assumptions, the overall class learning rate is proportional to  $K \times U$ . (Typical applications of peer instruction are followed by instructor review or clarification and do not fulfill this criterion.)

Up to this point, we have assumed that the knowledge with which each new nugget must be connected lies in the test domain, is generally implicit in constructivist theories. This could be verified by research showing that an independent (of the pretest) test of domain knowledge correlates with normalized gain, but we do not know of any such research.

On the other hand, it has been found that normalized gain scores correlate with the Lawson test<sup>25</sup> and also with the SAT Math scores.<sup>26</sup> This implies that the learning rate is proportional to prior knowledge *external* to the test domain,  $K_{external}$ , rather than with  $K_T(t)$ ,

$$dU_T/dt = -\alpha_{conn \ external} U_T(t) K_{external}, \quad (8)$$

with the solution

$$K_T(t) = 1 - (1 - K_{T0}) \exp[-\alpha_{conn \ external} K_{external} t], \quad (9)$$

$$g(K_{T0}) = 1 - \exp[-\alpha_{conn \ external} K_{external} t_{instr}]. \quad (10)$$

Since  $K_{external}$  is external to the testing (and instruction) domain, it does not increase with instruction. Thus, the func-

tional dependence on  $t$  is the same as the pure memory model, except with a learning probability,  $\alpha_{memory}$ , that depends on  $K_{external}$ . If  $K_{external}$  is known (e.g., from administration of a test on the enabling external knowledge), that value can or should be substituted for  $K_{external}$  in Eq. (8). Lacking such information, it is reasonable to take  $K_{external} = K_{T0}$ . This assumption may be justified either by assuming that, prior to instruction, internal and external knowledge are both proportional to the “general knowledge” of the student or by assuming that the students have had equal instruction in the domain previously and that those with higher relevant external knowledge have learned more and have a higher  $K_{T0}$  in proportion to  $K_{external}$ .

### C. Connectedness model

The pure memory and simple connected models represent ideal cases; obviously real learning can involve some learning of each type. (Indeed, some claim that tabula rasa involves some connected learning once the blank slate starts to fill or that constructivist learning starts with memory.) Therefore, we now introduce a model that interpolates between (and even beyond) the two pure models, being pure memory for  $\beta=0$  and simple connected for  $\beta=1$ . We call  $\beta$  the *connectedness parameter* because it is a measure of the degree of connected learning. In this model, a fraction  $\beta$  of the learning is connected, while the remaining  $(1-\beta)$  of the learning is assumed to be pure memory,

$$dU_T/dt = -U_T(t) [\alpha_{connected} \beta K(t) + \alpha_{memory} (1 - \beta)], \quad (11)$$

The knowledge  $K(t)$  connected to in Eq. (11) can be either  $K_T(t)$  (internal) or  $K_{external}$  (external to the test domain) leading to two different results for student post-test score and normalized gain,

$$K_T^{internal}(t) = 1 - \frac{(1 - K_{T0}) [\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta]}{(1 - K_{T0}) \alpha_{connected} \beta + [\alpha_{memory} (1 - \beta) + K_{T0} \alpha_{connected} \beta] \exp\{[\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta] t\}}, \quad (12)$$

$$g^{internal}(K_{T0}) = \frac{\exp\{[\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta] t_{instr}\} - 1}{\frac{1 - K_{T0}}{\alpha_{memory} (1 - \beta)} + \exp\{[\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta] t_{instr}\} + \frac{\alpha_{memory} (1 - \beta)}{\alpha_{connected} \beta} + K_{T0}}, \quad (13)$$

$$K_T^{external}(t) = 1 - (1 - K_{T0}) \exp\{-[\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta K_{external}] t\}, \quad (14)$$

$$g^{external}(K_{T0}) = 1 - \exp\{-[\alpha_{memory} (1 - \beta) + \alpha_{connected} \beta K_{external}] t_{instr}\}. \quad (15)$$

### D. Tutoring model

By our definition, a perfect tutor provides instruction based on complete knowledge of what the tutee does and does not know. The key point is that the tutor need not waste time re-instructing what the tutee already knows (as must a classroom instructor) and therefore can impart knowledge at the student’s maximum assimilation rate  $k_a$  (perhaps slightly

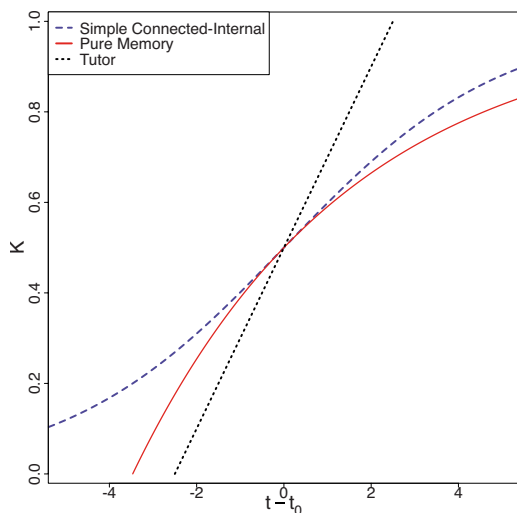


FIG. 1. (Color) Predictions of knowledge  $K(t)$  from our three “pure” models. We have assumed that  $k_a = \alpha_{memory} = \alpha_{connected}/2$ , which matches the learning rates of simple connected and pure memory models at  $t=0$  (where  $K_{T0}=0.5$ ). The learning rate is proportional to the slope. Thus, the tutoring model educates fastest everywhere, even for  $t < 0$  (i.e., before the pretest). The simple connected-external model (not shown) has the same time dependence as the pure memory model (although it differently scales depending on  $K_{external}$ ).

reduced because the tutor has to spend some fraction of time ascertaining what the tutee knows). The assumption that the tutee’s learning rate is independent of  $K(t)$  and  $U(t)$  is justified even if the student actually learns connectedly because the tutor can help the tutee connect each new knowledge nugget with something the tutor knows that the tutee already knows. Therefore, this model is characterized by a uniform rate of learning,

$$dU_T/dt = -k_a, \tag{16}$$

with the solution

$$K_T(t) = k_a(t - t_0), \tag{17}$$

implying

$$g(K_{T0}) = (k_a t) / [1 - K_{T0}]. \tag{18}$$

According to this model, a student can learn a finite test domain in a finite time; therefore, we additionally restrict  $K_T \leq 1$ .

This model should apply to self-tutoring on a task such as learning the capital cities of all the United States where the student can readily determine whether he knows each piece of knowledge. If the student learns what he does not already know connectedly, this model does not apply (because, unlike the perfect tutor, he does not know what to connect the new knowledge to) (Fig. 1).

### E. Discussion of models

Mathematical details aside, the differences between the models obviously follow from their motivating philosophies.

In tabula rasa as idealized in the pure memory model, the student is simply memorizing some of the presented information without connection to what he already knows; his absolute learning rate slows as his knowledge increases because more and more of the instruction addresses what he already knows. Constructivist theories insist that new knowledge must be associated with existing knowledge and, therefore, students ought to learn relatively faster when they have more knowledge—this is expressed in the connectedness model by a higher connectedness parameter implying faster learning when there is more prior knowledge to connect with. The ideal tutor can direct the instruction so that it always addresses what the student does not yet know (and if necessary connect it with what the tutor already knows the student does know), hence tutoring results in the highest learning rate, with its most significant advantage being when the student has learned most of the knowledge in the domain.

Our first three models are connected with usual prepost test analysis by noting that the class-average normalized gain, equal to  $1 - \exp[-\alpha_{memory} t_{instr}]$ , is the usual measure of the overall amount of learning. The connectedness coefficient  $\beta$  parametrizes the dependence of learning (transfer) on prior knowledge as measured by the systematic dependence of  $g$  on prescore. If changes in instruction result in a decrease in the measured connectedness parameter, this is an indication that students are now learning using more memory rather than some process of relating the instructional information to what they already know.

The tutoring model indicates that perfect tutors outperform other instructional methods by a higher margin at high prescores since conventional instruction wastes much time teaching what is already known. Hence, tutoring should be by far the best instructional method for those with a high initial level of knowledge. The use of one-on-one apprentice-based learning for graduate research students, craftspeople, and complex skills such as hunting and gathering seems to acknowledge this.

### III. DATA

In order to see whether our models can represent reality and allow us to draw conclusions about it, we apply them to concept test scores for students in single classes, a sharp test of the models because all students receive the same instruction. The fitted data are for large classes and may be considered “high quality” from the standpoint of educational studies—even binned into six to ten bins, relative errors are a few percent, and fits to one data set give parameters excluded for another with astronomically small  $p$  values (e.g.,  $p < 10^{-16}$ ). They therefore provide a stringent test of the models’ ability to fit high-quality data.

The MIT data are from two versions of the largest required introductory calculus-based mechanics class ( $N \sim 400$ ) at MIT. (This class is not taken by the most skillful  $\sim 35\%$  or the least skillful  $\sim 10\%$  of freshmen, who select other options.) The 2003 MIT class used the traditional lecture-recitation format, while the 2005 MIT class used studio format. The studio format utilized both peer instruction (five to ten concept questions per week) and group problem



solving of typical multipart problems. MASTERINGPHYSICS<sup>27</sup> which has been shown to impart some conceptual knowledge<sup>28</sup> was used in both MIT classes, but somewhat more extensively in 2003 than 2005.

The University of Minnesota (UMn) data are from several successive years of the corresponding course at the UMn and were given to us by Henderson *et al.*<sup>29</sup> These classes utilized group solving of context-rich problems<sup>5,30,31</sup> to impart conceptual knowledge. UMn used the FCI<sup>32</sup> and MIT used the MBT,<sup>33</sup> both concept tests composed by members of the Arizona State University educational research group. The MBT was given to MIT students because it (but not the FCI) covers energy and momentum, contains some quantitative problems, and is designed for students with prior formal training in mechanics which almost all MIT students have.

With statistical certainty, the MIT data imply that  $\beta$  is close to 0. Taking both classes together suggests that it is significantly below 0, i.e., that normalized gain slightly decreased with increasing prescore. This would be the expected deviation from pure memory-based learning if the “pass/no record” grading of this required course encouraged students with high prescores to exert less effort and/or students with low prescores to exert more. We have independent evidence of such behaviors (e.g., top students doing less homework near the end of the term and bottom students going to special tutorials). Granting this systematic effect, we conclude that the MIT data are consistent with the pure memory model, suggesting that memory-based learning predominates.

The fit to the UMn data, on the other hand, has  $\beta = 0.45 \pm 0.04$ , which implies a very significant amount of connected learning. Although the connectedness-external model fit of the UMn data lowered the value of the reduced chi-squared from 1.54 to 1.19 relative to the connectedness-internal model, we do not regard this improvement as significant since different bin sizes have been observed to change the reduced chi-squared by up to  $\pm \frac{1}{2}$  or so. (Furthermore, the best fits using these two models differ only in the outermost bins where systematic effects are largest, see Fig. 2.)

We suggest that the highly significant difference in the connectedness parameter between MIT and UMn does indeed reflect the preferential use of memory vs constructed knowledge in the transfer of information from instruction to assessment. At MIT, the instruction methods, peer instruction<sup>34</sup> or MASTERINGPHYSICS,<sup>27</sup> use both the same representation and context (multiple choice questions with primarily textual or graphical choices) as the test. Such “teaching like the test” may well encourage students to respond to the post-test by referring to a similar remembered concept question rather than by reasoning from any constructed knowledge. At UMn, the instruction is in a different representation and context than the test-group solving of context-rich problems. Performance on the test therefore requires “far transfer” (of knowledge), which involves constructed knowledge.

The result that  $\beta \sim \frac{1}{2}$  (not 1) at UMn indicates that their students use considerable memory-based learning and do not learn and transfer knowledge to Newtonian concept tests fully connectedly. It may be that students in introductory courses often lack the requisite knowledge to construct

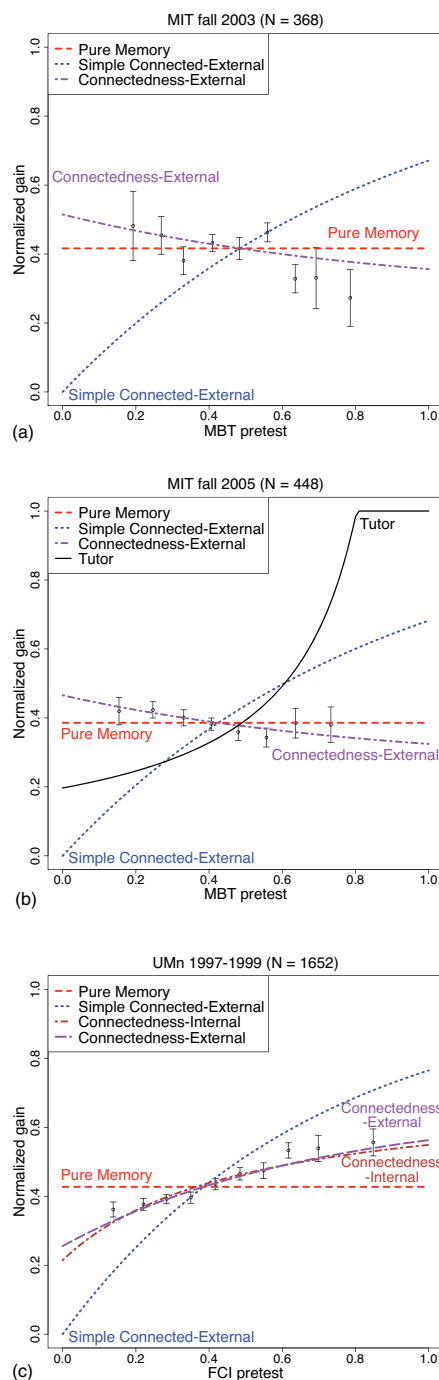


FIG. 2. (Color) Fits of pure memory, connectedness, and simple connected-external models to the normalized gain of (A) the MIT 2003 class, (B) MIT 2005 class, and (C) UMn 1997-1999 classes. The tutoring model is shown in (B) to illustrate its positive curvature in a normalized gain plot. For these plots, each data set was divided into bins of equal spacing except for two inclusive bins at the extremes of the pretest range (hence the slightly wider separation of the edge points), and the average and STD of the mean of the normalized gains of the students in each bin were plotted against their average pretest score.  $\alpha_{memory}^{I_{instr}}$  and  $\alpha_{connted}^{I_{instr}}$  were free parameters for the pure memory and the simple connected model fit. In the connectedness model fit,  $\beta$  was determined by extrapolating between pure memory and the simple connected model. Results of fits are in Table I.

TABLE I. Fit of models to three data sets described in text. Pure memory and simple connected-external models were first fit to determine the products  $\alpha_{memory}^{t_{instr}}$  and  $\alpha_{connected}^{t_{instr}}$ . The probability that the data are consistent with each of the pure models is shown as  $p_{memory}$  and  $p_{connected}$ , respectively. Then, the interpolation parameter  $\beta$  was varied to produce the lowest reduced chi-squared,  $\chi^2_\nu$ .

Class	Model	$\alpha_{memory}^{t_{instr}}$	$p_{memory}$	$\alpha_{connected}^{t_{instr}}$	$p_{connected}$	$\beta$	$\chi^2_\nu$
MIT 2003	Connectedness-external	$0.53 \pm 0.02$	0.07	$1.11 \pm 0.05$	$1.02 \times 10^{-6}$	$-0.24 \pm 0.16$	1.91
MIT 2005	Connectedness-external	$0.49 \pm 0.02$	0.01	$1.15 \pm 0.04$	$8.34 \times 10^{-19}$	$-0.21 \pm 0.09$	0.43
UMn	Connectedness-internal	$0.56 \pm 0.01$	$3.93 \times 10^{-17}$	$1.45 \pm 0.03$	$3.93 \times 10^{-17}$	$0.50 \pm 0.06$	1.54
UMn	Connectedness-external	$0.56 \pm 0.01$	$6.22 \times 10^{-16}$	$1.45 \pm 0.03$	$1.78 \times 10^{-33}$	$0.40 \pm 0.05$	1.19

knowledge from a presented nugget but find that remembering it is a better strategy than ignoring it. Remembering something or reasoning by superficial analogy from something remembered is a valuable strategy to enhance performance on a timed test (the concept tests allow approximately 2 min per question).

We investigated whether the MIT—UMn differences might be caused by the differences between the MBT and FCI tests used. To check this possibility, we transformed the MBT scores into FCI scores by using several functions that fit the data in Wells *et al.*<sup>35</sup> for groups of students given both tests at nearly the same time. A typical transformation function to obtain FCI-equivalent normalized gain from MBT data is

$$\begin{aligned} \text{normalized FCI} &= \text{normalized MBT} \\ &+ 0.11 \quad (\text{when normalized MBT} \leq 0.8), \\ \text{normalized FCI} &= 6.40 \times \text{normalized MBT} / \\ &(1 + \text{normalized MBT}^2) \\ &- 2.21 \quad (\text{when normalized MBT} > 0.8). \end{aligned} \tag{19}$$

Different transformation functions gave different effective FCI scores for the MIT classes. Fits to these gave  $\beta$ 's between  $-0.07$  and  $-0.42$ , but the average  $\beta$  was very close to the values reported above. In addition, we performed our analysis by using only the 16 questions on the MBT that have multiple choice text-based answers like all those on the FCI (i.e., excluding MBT questions requiring calculation). It gave results for  $\beta$  well within error of those obtained based on the entire test.

There can be many systematic effects on  $\beta$  when comparing classes at different institutions taught with different pedagogies by different professors and using different assessments. We have mentioned student motivation due to pass/no record grading. It would seem that the professor could influence  $\beta$  by aiming the instruction at students with low vs high prescores.

We suggest that year to year variations in  $\alpha$  or  $\beta$  would be probative as to the overall amount and type of learning associated with different instructors or pedagogies at the same institution, with the same subject and using the same test,

even if systematic effects due to the particular test or the particular institutional circumstances do exert a large systematic effect on the differences between classes at different institutions. To establish that the connectedness parameter has a meaningful interpretation across institutions will require an extensive study of real data. Nevertheless, it appears that a complete characterization of prepost data should involve fitting both  $\alpha$  and  $\beta$ .

#### IV. DISCUSSION

The various models in this paper enable us to fit data like Hake's and clearly give connectedness parameter  $\sim 0$  for his data. Interpreting this in terms of the models suggests predominance of memory-based learning for both traditional and interactive classes. Perhaps, the lower normalized gain of the traditional classes results simply from the lack of student engagement typical of such settings. These models apply more cleanly to variations with prescore within individual classes, and our success in fitting the high-quality MIT and UMn data should inspire others to make similar fits to their data from prepost testing with intervening instruction and to try to draw conclusions about the type of learning or transfer that they are observing. Anticipating and hoping to encourage this, we now discuss some caveats about this work, some conclusions, and some challenges that it raises, especially with regard to the models and their implications.

##### A. Caveats

“Learning” is the absorption or incorporation of new knowledge into the mind and cannot be measured without some sort of before and after assessment of the learner. When the student's learning is measured by some assessment, our models really involve knowledge transfer from instruction to assessment. For example, the type of learning displayed by MIT students might be significantly connected if assessed by using transfer to context rich problems. Similarly, the UMn students might display more pure memory-based learning if assessed by using context rich problems similar to those used for instruction.

Once we have distilled key ideas from a typical learning theory into mathematical assumptions, those assumptions completely characterize the content of the model. A fit of the resulting model to the data indicates only that those math-

emathical assumptions, but not all the ideas of the theory motivating the model, can describe the dependence of score improvement of students with different prior knowledge. (Moreover, it may be possible to derive the same model from seemingly different assumptions.) Obviously, the connectedness parameter measures the advantage of a student with more prior knowledge in transferring instruction into post-test improvement. Our data illustrate one case that seems to accord with expectations based on the learning theories that motivate the models.

Data fit by our models are subject to many systematic errors. While most of those discussed above do not apply to comparisons made for similar classes with the same assessment, the effects of assuming a distribution of individual sticking coefficients, testing error, or varying difficulties of knowledge nuggets should be numerically investigated (although averages like this rarely make huge differences in fit value or slope, which correspond roughly to  $\alpha$  and  $\beta$ ). Obviously, inclusion of a few very easy or very hard items on the tests will cause a systematic tilt of the normalized gain [Eq. (1)] that could affect the extracted value for  $\beta$ . Still, when using the same assessment instrument, especially at the same institution with intervening changes in instructional design, it seems reasonable to argue that the overall normalized gain and connectedness parameter are related to the overall efficacy of instruction and its degree of constructedness.

### B. Conclusions

Learning theories generally classed as tabula rasa, constructivist, and tutoring assign different roles to the prior knowledge of those instructed. These motivate our different mathematical learning models.

The connectedness model is shown to fit high-quality prepost test data almost within error across a factor of 5 variation of both initial  $K_{T0}$  and  $U_{T0}$ , fitting the data even when both pure models are excluded by astronomical  $p$  values. It seems likely to be a useful parametrization for the many sets of observed data with correlation between normalized gain and prescore.<sup>25</sup>

The fact that our two-parameter model fits excellent data so well shows that more sophisticated mathematical learning models (e.g., representing additional features of constructivist hypotheses) are unlikely to fit the data significantly better and will likely be less preferred as explanations for high-quality prepost test data on the basis of Occam's razor. More detailed assessments of students (e.g., other types of assessment) seem necessary to justify models with more detailed assumptions than the simple ones made in our models.

### C. Challenges and open questions

A central question is raised by our work together with Hake's data and the many studies showing quite constant normalized gain vs pretest score. Why are these data consistent with the pure memory learning model [Eq. (4)], suggesting memory-based learning, when memory-based theories of learning have been largely supplanted by constructivist theories? (For example, if constructivist learning is, in fact, occurring, why do students with more domain knowledge and

therefore higher pretest scores not learn faster and therefore have higher normalized gain as predicted by either our connected model or the self-tutoring model? Is it possible that peer instruction using questions similar to those on the concept tests allows students to transfer knowledge to those tests mainly by memory or simple analogies?) A less important, but still vexing question is as follows: "How is it possible that many university classes have shown a significant positive correlation of gain and pretest<sup>36</sup> while Hake's averages of classes show essentially none?"

Turning from the connectedness models to the tutoring model, it would be interesting to see if data on student learning of vocabulary words follow the tutoring model as suggested earlier. In addition, learning data from the growing number of computer-based personal tutors should approach this model as the tutors approach our ideal of perfection. Several authors<sup>26,37,38</sup> have suggested that lack of particular identifiable and measurable skills influences the normalized gain. Our connectedness model should fit data involving this effect if  $K(t)$  in Eq. (11) were replaced by  $K_{external}$ , the measured external knowledge. We believe that the key challenge (especially to those who believe in using outcomes of instruction for some useful purpose) is to investigate whether the models here can be used to provide insight into data on learning. Hake's paper showing that the normalized gain [ $g = 1 - \exp(-\alpha t_{instr})$ ] is primarily determined by the mode of instruction offers some hope that the connectedness parameter  $\beta$  will also prove to be a parameter that is primarily sensitive to the type of learning and transfer. Showing this will require a study of similar magnitude to his, however.

### V. SUMMARY

Our models make specific mathematical predictions that are based on simple assumptions motivated by different theories of learning. The connectedness model introduces a connectedness parameter that indicates the degree to which learning depends on the student's prior knowledge. Based on the ideas underlying the models, this parameter indicates the relative amount of memorization vs associative transfer from instruction to assessment. Existing data are well fit by this model, sharply determining this parameter. When validated by comparison with more examples (and possibly generalized to account for more effects), such models may well allow diagnosis of *how* students are learning from the repeated measurements of *how much students know* that are already being made for other purposes. We urge researchers who publish prepost test data to fit it by using the connectedness model and to report both  $\alpha$  and  $\beta$ .

### ACKNOWLEDGMENTS

We are grateful to C. Henderson for providing us with the UMn data. We are grateful to E. Mazur and N. Lasry for valuable discussions and comments on the manuscript. D.E.P. and Y.-J.L. acknowledge support from MIT; L.B. acknowledges support from OSU. This work was supported in part by NSF Grant No. 0457451.

- <sup>1</sup>R. R. Hake, Interactive-engagement vs. traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- <sup>2</sup>MIT used the mechanics baseline test (MBT) (Ref. 33), whereas UMN used the force concept inventory (FCI) (Ref. 32), as did Hake (Ref. 1).
- <sup>3</sup>J. Locke, *An Essay Concerning Human Understanding* (Clarendon, Oxford, 1690).
- <sup>4</sup>L. Resnick and A. Collins, Cognition and Learning, in *The International Encyclopedia of Educational Technology*, edited by T. Plomp and D. Ely (Pergamon, Oxford, 1996), p. 48.
- <sup>5</sup>R. Mayer, Learners as information processors: Legacies and limitations of educational psychology's second metaphor, *Educ. Psychol.* **31**, 151 (1996).
- <sup>6</sup>D. Jonassen, Objectivism vs. constructivism: Do we need a new paradigm? *Educ. Technol. Res. Dev.* **39**, 5 (1991).
- <sup>7</sup>E. von Glasersfeld, An Introduction to Radical Constructivism, *The Invented Reality*, edited by P. Watzlawick (Norton, New York, 1984), p. 17.
- <sup>8</sup>H. Gardner, The academic community must not shun the debate over how to set national educational goal, *Chronicle Higher Educ.* **36**, A53 (1989).
- <sup>9</sup>R. R. Hake, Socratic pedagogy in the introductory physics laboratory, *Phys. Teach.* **30**, 546 (1992).
- <sup>10</sup>It is common to administer the same instrument for both, but the post-test for the previous year can serve as the pretest for the current year if suitably calibrated.
- <sup>11</sup>J. D. Bransford and D. L. Schwartz, Rethinking Transfer: A Simple Proposal with Multiple Implications, in *Review of Research in Education*, edited by Iran-Nejad and D. Pearson (American Educational Research Association, Washington, DC, 1999), p. 61.
- <sup>12</sup>This half-century-old pregain and/or postgain index was independently utilized by Hovland *et al.* who called it the "effectiveness index" (Ref. 13), Gery who called it the "gap-closing parameter" (Ref. 14), and Hake who called it the "normalized gain" (Ref. 1) as we do.
- <sup>13</sup>C. I. Hovland, A. A. Lumsdaine, and F. D. Sheffield, A Baseline for Measurement of Percentage Change, in *The Language of Social Research: A Reader in the Methodology of Social Research*, edited by P. F. Lazarsfeld and M. Rosenberg (Free, New York, 1955), p. 77.
- <sup>14</sup>F. W. Gery, Does Mathematics Matter?, in *Research Paper in Economic Education*, edited by A. Welch (Joint Council on Economic Education, New York, 1972), p. 142.
- <sup>15</sup>We assume that any knowledge element tests either known or unknown (right or wrong) thus sidestepping complications due to well-established alternative conceptions (Refs. 16–18) and mixed mental states (Refs. 19,20).
- <sup>16</sup>S. Panse, J. Ramadas, and A. Kumar, Alternative conceptions in Galilean relativity: Frames of reference, *Int. J. Sci. Educ.* **16**, 63 (1994).
- <sup>17</sup>A. I. Zietsman and P. W. Hewson, Effect of instruction using microcomputer simulations and conceptual change strategies on science learning, *J. Res. Sci. Teach.* **23**, 27 (1986).
- <sup>18</sup>P. Heller and F. N. Finley, Variable uses of alternative conceptions: A case study in current electricity, *J. Res. Sci. Teach.* **29**, 259 (1992).
- <sup>19</sup>Z. Hrepic, D. A. Zollmann, and S. Rebello, Identifying Students' Models of Sound Propagation, in *Physics Education Research Conference, Boise, 2002*, edited by S. Franklin, K. Cummings, and J. Marx (American Institute of Physics, New York, 2002), p. 59.
- <sup>20</sup>L. Bao and E. F. Redish, Model analysis: Representing and assessing the dynamics of student learning, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010103 (2006).
- <sup>21</sup>W. K. Estes, Toward a statistical learning of theory, *Psychol. Rev.* **101**, 282 (1994).
- <sup>22</sup>R. A. Rescorla and A. R. Wagner, A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement, in *Classical Conditioning II: Current Research and Theory*, edited by A. H. Black and W. F. Prokasy (Appleton-Century-Crofts, New York, 1972), p. 64.
- <sup>23</sup>R. C. Atkinson, G. H. Bower, and E. J. Crothers, *An Introduction to Mathematical Learning Theory* (Wiley, New York, 1965), Chaps. 1 and 2.
- <sup>24</sup>C. L. Hull, *A Behavior System* (Yale University Press, New Haven, 1952).
- <sup>25</sup>V. P. Coletta and J. A. Phillips, Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability, *Am. J. Phys.* **73**, 1172 (2005).
- <sup>26</sup>R. L. Hill, D. Grosnick, and D. Ober, Student preparation, aptitude, and performance in a first-semester algebra-based physics course, *Bull. Am. Astron. Soc.* **38**, 1235 (2007).
- <sup>27</sup>MASTERINGPHYSICS, Addison-Wesley, ([www.masteringphysics.com](http://www.masteringphysics.com)).
- <sup>28</sup>E.-S. Morote and D. E. Pritchard, Annual Meeting of the National Association for Research in Science Teaching, New Orleans, LA, 2002 (unpublished).
- <sup>29</sup>C. Henderson, P. Heller, and the UMN Physics Education Research Group, Contributed Talk, American Association of Physics Teachers Winter Meeting, Kissimmee, FL, 19 January 2000 (unpublished).
- <sup>30</sup>P. Heller, R. Keith, and S. Anderson, Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving, *Am. J. Phys.* **60**, 627 (1992).
- <sup>31</sup>P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- <sup>32</sup>D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, *Phys. Teach.* **30**, 141 (1992).
- <sup>33</sup>D. Hestenes and M. Wells, A mechanics baseline test, *Phys. Teach.* **30**, 159 (1992).
- <sup>34</sup>E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
- <sup>35</sup>M. Wells, D. Hestenes, and G. Swackhamer, A modeling method for high school physics instruction, *Am. J. Phys.* **63**, 606 (1995).
- <sup>36</sup>R. R. Hake, Physics Education Research Conference, Boise, Idaho, August 2002 (unpublished).
- <sup>37</sup>D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible 'hidden variable' in diagnostic pretest scores, *Am. J. Phys.* **70**, 1259 (2002).
- <sup>38</sup>R. K. Thornton, Uncommon Knowledge: Student Behavior Correlated to Conceptual Learning, in *International School of Physics Enrico Fermi Course CLVI, Varenna, Italy, 2004*, edited by E. F. Redish and M. Vicentini (IOS, Amsterdam, 2004).