Learning Curves in Health Professions Education

Martin V. Pusic, MD, PhD, Kathy Boutis, MD, MSc, Rose Hatala, MD, MSc, and David A. Cook, MD, MHPE

Abstract

Learning curves, which graphically show the relationship between learning effort and achievement, are common in published education research but are not often used in day-to-day educational activities. The purpose of this article is to describe the generation and analysis of learning curves and their applicability to health professions education. The authors argue that the time is right for a closer look at using learning curves—given their desirable properties—to inform both self-directed instruction by individuals and education management by instructors.

A typical learning curve is made up of a measure of learning (y-axis), a measure of effort (x-axis), and a mathematical linking function. At the individual level, learning curves make manifest a single person's progress towards competence including his/her rate of learning, the inflection point where learning becomes more effortful, and the remaining distance to mastery attainment. At the group level, overlaid learning curves show the full variation of a group of learners' paths through a given learning domain. Specifically, they make overt the difference between time-based and competency-based approaches to instruction. Additionally, instructors can use learning curve information to more accurately target educational resources to those who most require them.

The learning curve approach requires a fine-grained collection of data that will not be possible in all educational settings; however, the increased use of an assessment paradigm that explicitly includes effort and its link to individual achievement could result in increased learner engagement and more effective instructional design.

Learning curves graphically show the relationship between learning effort (e.g., repetitions or time spent; see also "Validity of the x-axis: Measurement of effort, experience, or practice" below) and the resultant learning outcomes.[1–3] The concept has become so ubiquitous that educators often use the term when describing learning processes (e.g., "a steep learning curve") without necessarily referring to a specific figure. Underlying the learning curve and its text descriptors are fundamental psychological truths that have important educational implications: Practice improves performance, and good practice improves performance even more; sufficient practice leads to high levels of achievement; and the most dramatic learning occurs early in the learning process.[2,4]

Please see the end of this article for information about the authors.

Correspondence should be addressed to Martin V. Pusic, Division of Education Quality and Analytics, Institute for Innovations in Medical Education, NYU School of Medicine, 545 First Ave., Suite 6P, New York, NY 10016; telephone: (212) 263-2053; e-mail: martin.pusic@nyumc.org.

Acad Med. 2015;90:1034–1042.
First published online March 20, 2015
doi: 10.1097/ACM.0000000000000681

Supplemental digital content for this article is available at http://links.lww.com/ACADMED/A266 and http://links.lww.com/ACADMED/A267.

The concept of the learning curve holds potential value far beyond a narrow psychological application. For a number of reasons, the time is right for educators to take a fresh look at the use of learning curves in the health professions. First, and perhaps most compelling, the recent emphasis by the Accreditation Council for Graduate Medical Education (ACGME) on pairing competency assessments with developmental milestones naturally favors individualized competency-based metrics,[5] potentially including learning curves. Second, learning curves aptly represent the learning acquired through deliberate practice, which is increasingly recognized as an effective instructional strategy.[6] Third, the longitudinal collection of increasingly granular, learner-specific information in computer databases facilitates tracking learner growth over time, a process that is naturally represented as a learning curve.[7] Finally, multilevel data modeling, an analysis technique that enables quantitative analysis of learning curves, has graduated from a specialized research technique to one accessible to educators.[8]

The purpose of this article is to enable health professions educators to enhance their teaching, assessment, and research activities by using learning curves more effectively. We first describe the anatomy of a learning curve and its useful properties for educators. We then discuss validity, which is an important consideration for all assessments but is especially complicated for learning curves as they entail three separate components, each of which requires separate consideration. We subsequently describe how to graph learning curves, followed by an outline of an approach to their quantitative and qualitative evaluation or analysis. We discuss possible health professions applications for learning curves, highlighting their usefulness in instruction and education research. Finally, we end with a brief review of potential pitfalls.

Anatomy of a Learning Curve

We show a generic learning curve in Figure 1. In general, the relationship between learning (or performance) and effort (or, in Figure 1, deliberate practice) is not linear. Classically, the learning curve can take the form of an ogive or S-shape indicating that the learning rate (or slope) varies as the location of the subject (hereafter "learner") changes during the learning process.[1,3,9] As an example, consider learning to interpret a pediatric ankle radiograph wherein the task is to look for a possible fracture.[1]

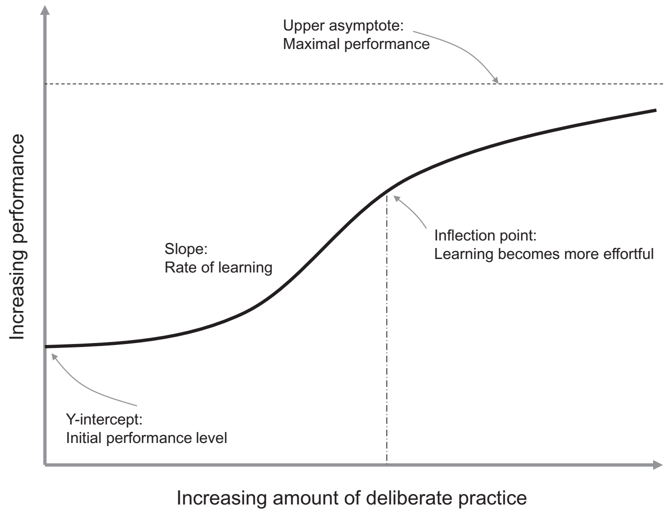

Figure 1 Generic learning curve showing key properties, including, on the y-axis, increasing performance (or learning), and, on the x-axis, increasing amount of deliberate practice (or effort). Adapted from Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. Acad Med. 2011;86:731–736.

Because even a layperson can interpret some radiographs correctly, baseline performance (indicated by the y-intercept) will usually not start at zero. With practice, the learner might typically show a "latent" phase during which performance on cases hardly improves while the learner becomes familiar with the elements of the domain ("Does this soft-tissue swelling count?" "Is that a growth plate or a fracture?"). Once the learner has learned the basics, his or her performance usually improves rapidly as evidenced by an acute upward slope in the learning curve. During this growth phase, learning efficiency is maximal; that is, the learner shows significant gains in learning with each repetition. However, the rate of learning eventually slows, causing an inflection, because not all to-be-learned aspects of a task improve performance equally. In other words, some of the tasks are not as easy (or hard) to learn.[4] In the ankle radiograph example, some fractures are more rare, are harder to perceive, or are inherently more difficult to understand, so it takes more repetitions to learn them. The curve gradually approaches an asymptote that represents the maximal performance achievable in the given learning context.[10]

Learning curves can be created for an individual learner or averaged together to estimate key features of learning performances among a group of learners.[1] We emphasize that separate considerations apply to the interpretation of individual curves as opposed to group curves, as we discuss below in the sections entitled Graphing Learning Curves and Analysis of Learning Curves.

Evidence Supporting the Validity of a Learning Curve

For learning curves to be useful assessment tools, educators must be able to construct a validity argument to support the use of a learning curve in a particular assessment or learning context.[11–13] However, making a validity argument for a learning curve is complicated by the fact that, in essence, three separate validity arguments must be made: one each for (1) the measure of learning (y-axis), (2) the measure of effort (x-axis), and (3) the measure of how they are related (linking function). In this section, we will consider each of these in turn.

Validity of the y-axis: Measurement of learning

For a learning curve to show the relationship between learning and effort, a valid, reliable, and responsive measure of learning is necessary.[13] For example, Figure 2 shows the results of progress tests given twice annually to Dutch medical students at Maastricht University.[14] We can see that over the six years of medical school, students progressed from a low score of approximately 10% in their first year and gained approximately 10% each year in their scores until at graduation they averaged 70%. To model the learning across the entire medical school curriculum, the educators needed an outcome measure that, in equal increments, spanned the knowledge gain seen in six years of instruction.

Current thinking on the validity of assessment measures invokes a paradigm in which validity is viewed as a hypothesis. Investigators collect evidence to support or refute the hypothesis that the measure is a valid reflection of desired learning (the so-called "construct" in question).[11–13] The evidence can derive from multiple sources.[11–13] *Content* evidence typically comprises data showing rigorous test development steps.[12,15] Evidence of *relations with other* variables shows how data relate to other facts and figures, such as other independently generated scores.[12] *Response process* evidence explores the "fit between the construct and the detailed nature of performance [and thought processes] … actually engaged in"[12,15]; in other words, did the radiology learner systematically examine all of the radiograph images in their entirety, or did she skip images or regions in a way that would not be tolerated in actual clinical practice?

Although these previous three types of evidence are not considerably different for the y-axis of a learning curve than for other assessments, the final two have particular considerations for learning curves. *Evidence of internal structure* comprises "classic" psychometric data, such as reliability, factor analysis, and item analysis. Learning curves can become distorted in the presence of floor effects (i.e., scores at the low end of a performance scale have decreased variability) or ceiling effects (i.e., high scores have decreased variability), resulting in a learning curve that appears vertically compressed along the y-axis (more shallow) than would be expected. *Consequences evidence* evaluates the intended or unintended effects of the act of assessment, such as effectively guiding decisions about remediation or student placements. If educators can use learning curves to effectively detect earlier which students are off-track, then this would constitute evidence of (beneficial) consequences.

Validity of the x-axis: Measurement of effort, experience, or practice

A number of measures can constitute the x-axis of a learning curve. Typically, the measure is a countable repetition or time measure or other activity that is known to be associated with the desired learning outcome; however, more general concepts such as "experience" or "practice" are also used, depending on the context

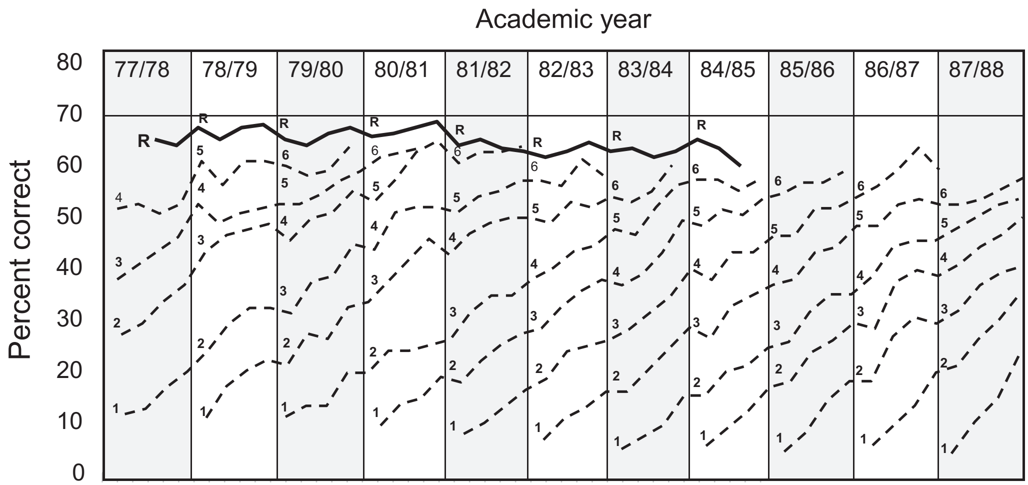

Figure 2 Maastricht progress tests. Average correct scores for each class or cohort per year (dashed lines) on successive progress tests from 1977 to 1988. Solid line (R) represents average scores of national reference groups of recently graduated doctors. Adapted with permission from Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach. 1996;18:103–110.

being modeled. For example, in Figure 2, the amount of time in medical school is used as the measure for the Dutch medical students. Using time alone, especially in a clinical setting where not all time "units" are equal, can result in problems or invalid results. In learning electrocardiogram (ECG) interpretation, a given resident will likely learn more about ECG interpretation during some rotations (e.g., cardiology) than in others (e.g., dermatology), and if months-in-training is used as the measure, the learning curve will show confounded changes in performance. Therefore, whenever possible, elapsed time should be replaced with measures that more accurately reflect the construct—namely, time actually spent in the learning activity being modeled (content evidence). For ECGs, this measure might be counts of the number of ECGs viewed.

Learning curves are frequently used to represent the process of deliberate practice, wherein the practice or repetition of a skill or maneuver always includes feedback; that is, each exposure comes with comments or reactions from an advanced instructor or guide.[2] This type of practice is effortful, incorporates expert-level feedback, and is available over an extended time period.[2] For deliberate practice, a valid measure will require both a means to accurately define and capture the number of repetitions and a way to ensure that feedback is available to the learner.

Throughout the rest of this article, we will refer to the x-axis as the *measure of learning effort*. We have chosen this admittedly general term in preference to *experience*, *practice*, *time*, or *repetition* because these other more specific terms can be context-bound and do not necessarily include the notion of active engagement (i.e., one might repeat an activity endlessly, but unless an effort is made to learn, this will not lead to learning or, therefore, a learning curve).

Measurement of association

Besides valid measures of learning and effort, a third element is required to create a meaningful learning curve: the linking function. This is defined as the mathematical equation that links the x-variable (effort) to the y-variable (learning or performance). The validity of the linking function depends on the fit of the psychometric relationship between the learning and effort and the extent to which potential confounders can be taken into account.[3] Linking functions will be described in detail below in the section entitled Analysis of Learning Curves.

Graphing Learning Curves

Having considered the validity of the data necessary to generate a learning curve, let us now turn to the task of generating a visual representation of learning curves. Key considerations include how to represent group means and individual variation, and how to represent multiple dimensions of a learning task.

First, group learning curves should represent both a group mean and some index of variability. Just as a simple group average score on a test is incomplete without a measure of variance (e.g., the standard deviation), a learning curve intended to reflect the performance of a group would ideally include both an average curve and some representation of variance. Figure 3 shows the previously published results of 18 postgraduate trainees learning to interpret ankle radiographs by serially interpreting 234 cases and receiving immediate feedback.[1] Specifically, Panel A shows the 18 individual learning curves, Panel B provides the average learning curve across all 18 learners, and Panel C illustrates the mean curve for the group, along with the 95% confidence intervals (CIs).[1] Note, however, that the CIs speak only to the range of *group-level* curves (i.e., if we repeated the study, 95% of the group-level average curves would be expected to fall within this range) and not the variability in individual curves that comprise those group averages. Finally, Figure 3, Panel D presents all 18 learning curves, superimposed. This representation shows that, although the overall curve can be seen to nicely follow a learning curve pattern, the paths of individual learners vary considerably. Note that an overall curve bounded by a 95% CI would not completely represent this variability because the CI speaks only to the range of possible average *group-level* curves and not the variability of the individuals within the group. Thus, to represent the learning curve variance, we recommend plotting overlaid individual learning curves, as in Figure 3, Panel D.

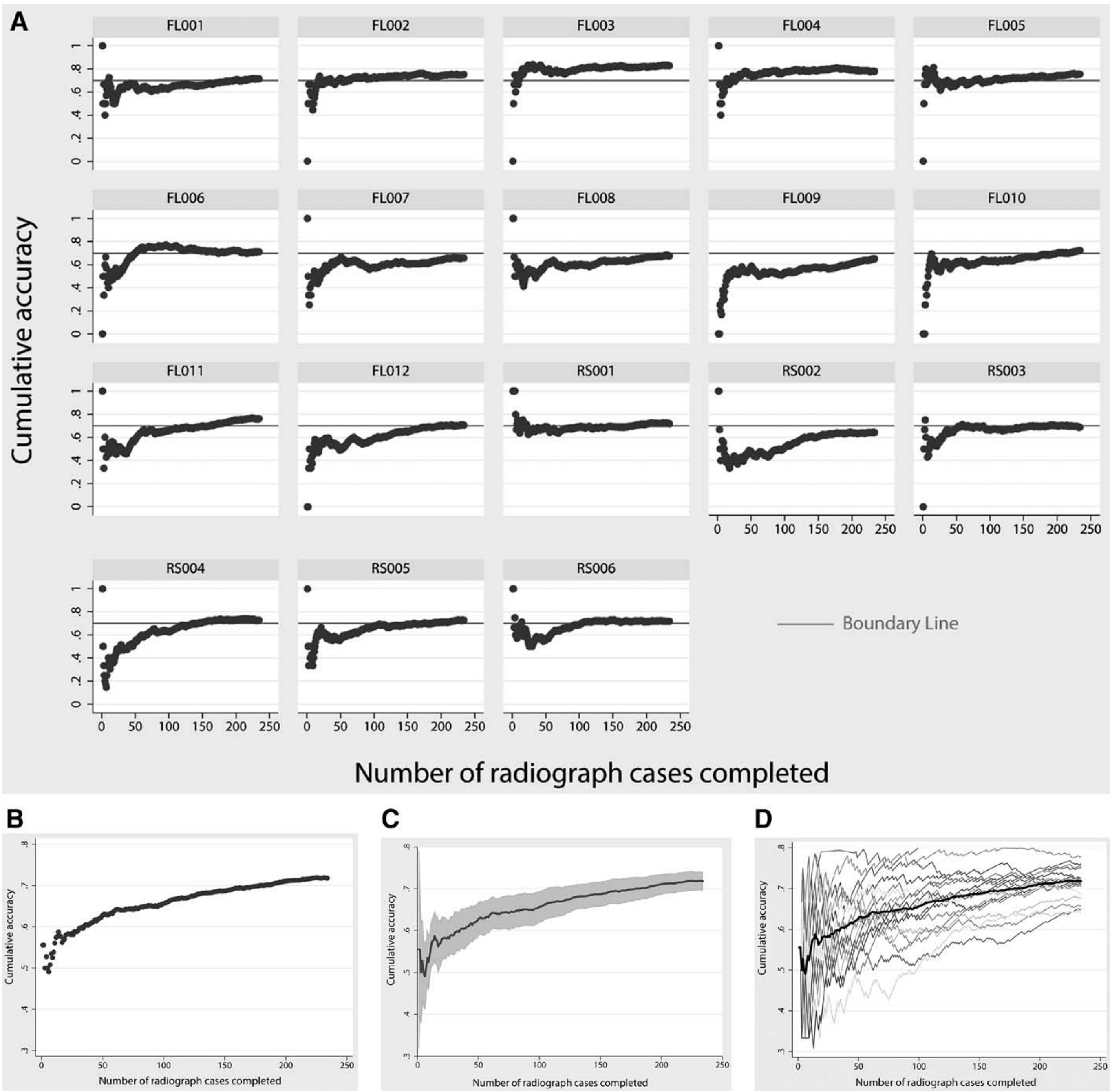

Figure 3 Learning curves for 18 residents interpreting pediatric ankle radiographs, with performance (cumulative accuracy) plotted against a measure of effort (the number of cases completed up to 234). Panel A shows the individual curves with the horizontal boundary line representing a competency threshold. Panels B and C show the group average curve, respectively, without and with the 95% confidence intervals (CIs). Panel D shows the individual curves overlaid. Variability between *individuals* is best conveyed with the overlaid curves and not a 95% CI which speaks only to the statistical uncertainty of the *group* average estimate. Panel A is adapted from Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. Acad Med. 2011;86:731–736.

For situations in which valid learning curves are available for several dimensions of a learning task (e.g., different domains of knowledge or skill, such as accuracy, speed, and confidence), stacking the learning curves can yield unique insights.[16] Figure 4 shows such a multidimensional learning curve. In this example, the x-axis and learner cohort are the same for all three stacked curves, but each y-axis represents a different dimension of the learning task.[17] This display allows a comparison of the respective developmental stages in the learning curve. For example, Figure 4 shows that speed continues to improve even after students in the group have slowed in their improvements in accuracy, whereas confidence varies considerably over time.

Analysis of Learning Curves

Learning curves can be computed for one individual or aggregated across many. In general, a descriptive analysis is used

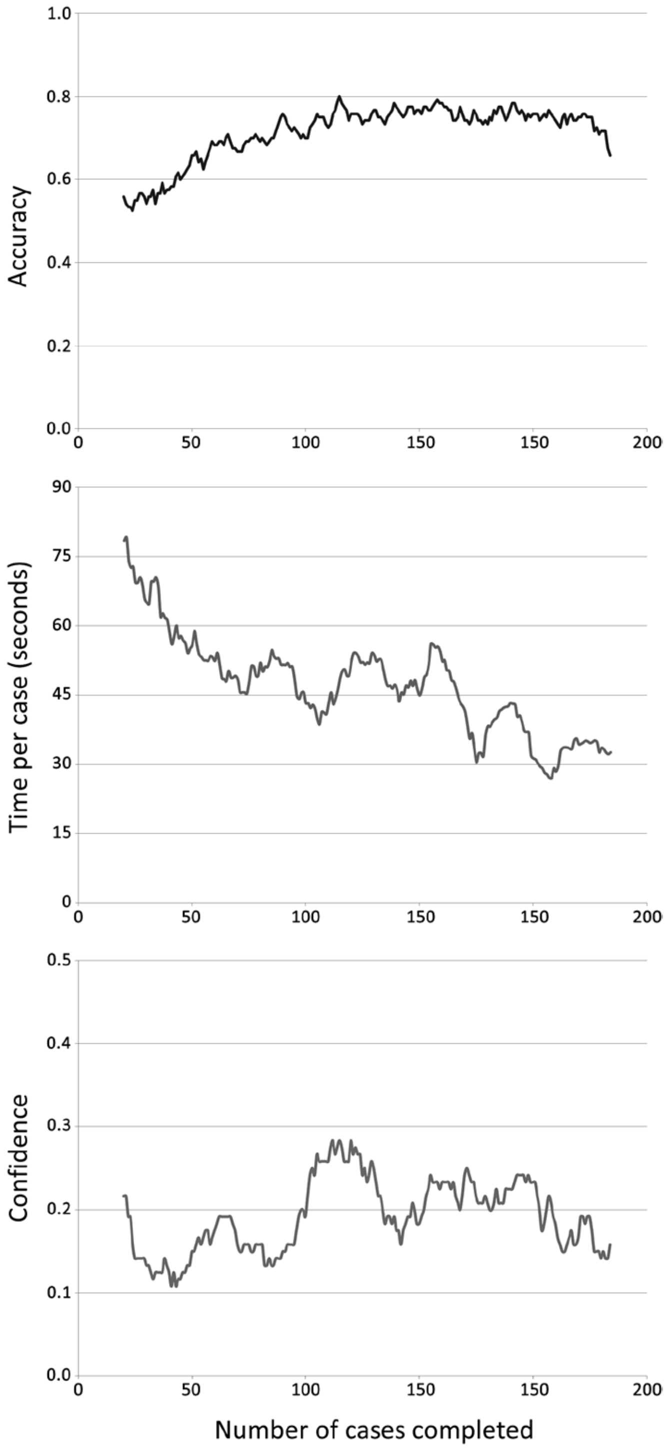

Figure 4 A multidimensional learning curve showing group averages for resident physicians practicing radiograph interpretation over 234 trials. Note that the participants plateau, in terms of accuracy, after about 150 trials but that their time-per-case continues to decrease through the last case. Their confidence in their ratings seems to start relatively high, decrease initially, and then rebound.

at the individual level, although some quantitative analyses are possible. At the group level, statistical modeling using a linking mathematical function is possible, assuming the group comprises a sufficient number of learners. Finally, multilevel modeling techniques hold promise in allowing group-level data to inform estimates for individual learners.

Individual-level analyses

The hallmark of individual-level learning curves is their marked variability. Consider again the learning curves in Panel A of Figure 3.[1] The wide heterogeneity of the learning paths taken by the trainees is immediately apparent. Each trainee started at different levels, progressed at different rates through the 234 cases, and finished with different abilities, albeit with decreased variance.

The shape of the individual curves provides interesting information about the learners' paths through the material—did they start off quickly and then fade (FL006) or stall and then fail to progress (RS006, RS001)? Were they not qualified or sufficiently ready to learn the material from the beginning (FL009, RS002)? Indeed, several learners had negative slopes to parts of their curves (FL004, FL006), which suggests that something was wrong with their learning experience. Learners who are proficient from the start might have flat or even decreasing learning curves (FL004). Each learning curve could be the basis for a conversation with the learner that would be richer than one based on a single numerical mean and standard deviation.

Group-level analyses

Averaging results across a group of learners results in a group-level learning curve (Figure 3, Panel B); however, the group average learning curve relationship holds for relatively few of the individual learners. The utility of the group learning curve lies in the information that it provides to the educator designing the intervention or assessment. Figure 3, Panel D shows the learning curves of all 18 radiology learners superimposed, along with the group-level learning curve. The group-level learning curve shows the type of negative exponential relationship (progressively decreasing slope) that we would expect on the basis of theories of deliberate practice.[10] If modeled using nonlinear regression techniques, we can derive estimates of parameters that would be useful to an educator who has been charged with optimizing the learning of a group of learners. *On average*, the learners start with an accuracy of 50%. From the slope, we see that they learn maximally over the first 100 repetitions, after which learning becomes less rapid. The asymptote for this learning intervention (an online case bank) appears to be at a cumulative accuracy of 72%.[1]

Linking equations

For the majority of situations in which a health professions educator would use a learning curve, the raw learning

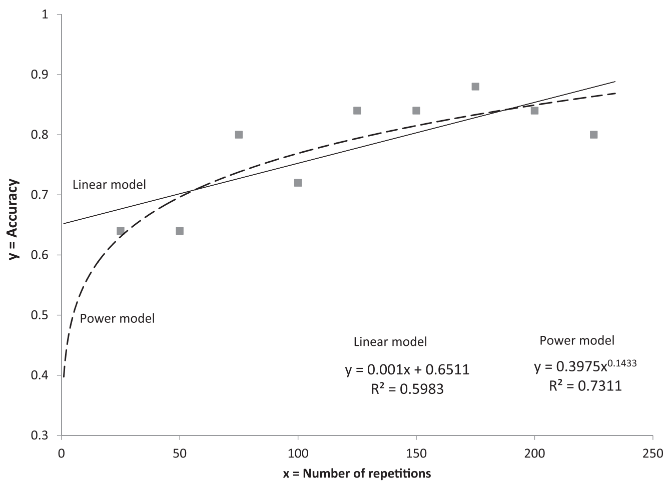

Figure 5 Modeling a single individual's learning curve. Accuracy versus number of repetitions for a single medical student deliberately practicing radiograph interpretation (data from Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. Acad Med. 2011;86:731–736). Each gray data point represents the student's score out of the 25 most recent repetitions. The authors have fitted both the linear (solid line) and power (dashed) regression models for illustration. This figure and its associated spreadsheet are included, respectively, as Supplemental Digital Figure 1 (available at http://links.lww.com/ACADMED/A266) and Interactive Supplemental CSV File 1 (available at http://links.lww.com/ACADMED/A267).

curves at the group and individual levels will suffice to guide learning as we have suggested; however, a number of potential mathematical relationships can be used to quantitatively describe the link between effort and learning in individual and/or group learning curves.[3] Choosing a linking function or model is a process akin to data modeling performed in biomedical research, in which researchers try candidate models in an iterative process guided by theory or measures of model "goodness of fit."[18] Although learning is rarely a direct linear function of the effort expended, the linear model (in which $y = a + bX$ and y is the measure of performance, a is the initial level of performance at time 0, b is the rate of learning, and X is the measure of effort) is an efficient first approximation. However, the linear model holds poorly at the higher end of the expertise scale (Figure 5). Instead, models based on the *power law* (in which $y = a + bX^{-c}$ and c is a measure of deceleration in learning as it progresses) are better able to represent the phenomenon of "diminishing returns" in which, at the expert end of a learning curve (approaching the asymptote), each invested unit of effort returns a smaller benefit in terms of performance improvement[2] (Figure 5).

Multilevel linear modeling holds promise as it allows researchers and educators to refine individual-level estimates based on the behavior of the group.[3,8,10] For example, knowing that medical students' learning differs from that of residents, as illustrated by the y-intercepts and slopes of their respective learning curves, allows us to refine predictions of how much deliberate practice is required to achieve a given level of competency for the different learner groups.[10] The greatest benefit of these types of quantitative "learning analytics" may lie in their capacity to use quantitative predictions to adapt and individualize learning within contexts that are suited to the generation of reproducible and valid learning curves.[19] To illustrate, in our radiology example, the position of the learner on the learning curve could be used to select the difficulty of the next case to be presented.

Boundary conditions: Illustrating cut scores or competency levels

A final consideration for graphing learning curves is the idea of the "boundary condition" or a threshold of learning that represents an important educational event. These are separate horizontal lines that represent criteria for the learning system such that, if the learning curve crosses the boundary line, an important educational result is signaled. Examples of these boundary conditions include a threshold of competency agreed upon by a consensus of educators (as shown in Figure 3, Panel A) or, in high-stakes learning, a level at which an insufficient learning rate results in an unacceptable number of failures over the time available for training. See Grigg and colleagues[20] and Holzhey and colleagues[21] for examples of cardiac surgeons' learning curves showing complication rates that eventually exceeded a predefined boundary, resulting in remediation.

Applications in Health Education

Learning curves as an assessment metric

In many respects, learning curves are an ideal assessment metric because, in a sense, they can demonstrate or make manifest the principle of assessment for learning.[22] If the learner receives the learning curve in real time (i.e., as he or she engages in the task), it can be sufficiently granular to demonstrate his or her path from novice to proficient (see Figure 6).[23] As mentioned, this information can be helpful at both the group and individual levels. The group information can guide the educator in terms of the overall nature of the relationship between effort and performance for the given task. It allows educators to answer group-level questions such as, How many repetitions must learners complete to achieve competence? Is there a latent phase? What is the maximum learning achievable? Does one instructional method lead to more efficient learning than another? How can I screen for learning problems as sensitively as possible?

The tight coupling of assessment with learning makes the learning curve ideal for formative assessment, providing reassurance for those who are on track and serving as an early detection system for those who are having difficulty. A ready example would be those learners in Figure 3 who have negatively sloped learning curves (FL004, FL006). However, the use of learning curves for summative assessment—enabling, for example, through the use of boundary lines, judgments of competence to advance to higher levels of practice—is less well understood and constitutes an area for future research.

Learning curves and competency frameworks

The Dreyfus and Dreyfus model of skill acquisition describes expertise development as a progression through several stages from a novice who is not

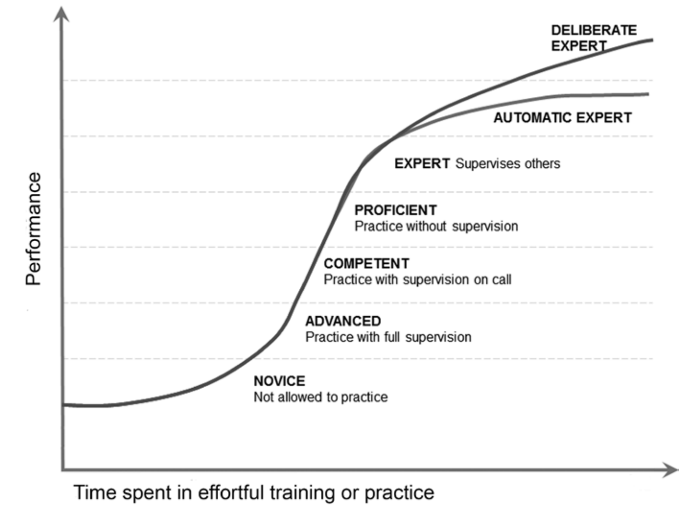

Figure 6 Models of expertise represented by learning curves. A conceptualization of the Dreyfus and Dreyfus model of skill acquisition extended to incorporate Ericsson's concept that some experts accept the stage of automaticity and stop improving, while others continue to seek out opportunities to improve in a deliberate manner, even though incremental improvements are hard-won at the expert stage. Adapted with permission from Kalet A, Pusic M. Defining and assessing competence. In: Kalet A, Chou CL, eds. Remediation in Medical Education. 1st ed. Boston, Mass: Springer; 2013.

allowed to practice on patients to a reflective expert who functions at the highest levels and continues to improve.[24] The Dreyfus and Dreyfus model[24] has been extended by Ericsson,[25] who points out that some experts "plateau" at an automatic level at which, absent deliberate effort to improve, successive repetitions do not lead to improvement in performance (the "automatic expert").

The Dreyfus and Dreyfus and the Ericsson models of expertise development can be useful in summarizing learning curves for assessing competency development. For example, the ACGME has "engaged the medical education community in articulating milestones of competency development"[5] wherein the milestones are observable behaviors in a sequential progression[26]—in essence, a learning curve. Figure 7 shows the difference between a time-based curriculum, through which all trainees spend the same amount of time in a program but graduate with different levels of competence; and a competency- or mastery-based curriculum, through which the final level of competence is standardized and the time needed to learn is allowed to vary. We believe that a deeper understanding of learning curves and their potential uses can inform the work of developing and extending these important "discourses on competence."[27]

Learning curves to support self-regulated learning

Theories of self-regulated learning focus on an iterative learning cycle, wherein learners think strategically about their learning, engage in the learning task, self-monitor, and then adjust their learning accordingly.[28] Learning curves may be helpful for self-regulated learning, especially in regard to self-monitoring and adjusting, as the visual representation of learning may encourage further reflection on the learning process.[9,29,30] Moreover, learning curves can facilitate self-regulation by showing learners their prior performance and anticipated future trajectory.[9] The effect of the explicit representation of the rate of learning and how that influences the learners' perception of challenge is an area of considerable interest.[9,31]

Finally, if performance continues to be tracked even after training ends, knowledge or skill may decay, a process that can be represented with a "forgetting curve." Experience curves join learning curves with forgetting curves to plot the full cycle of competency development and decay (Figure 8), thus aiding learners in recognizing when they require additional learning or refresher training.[32]

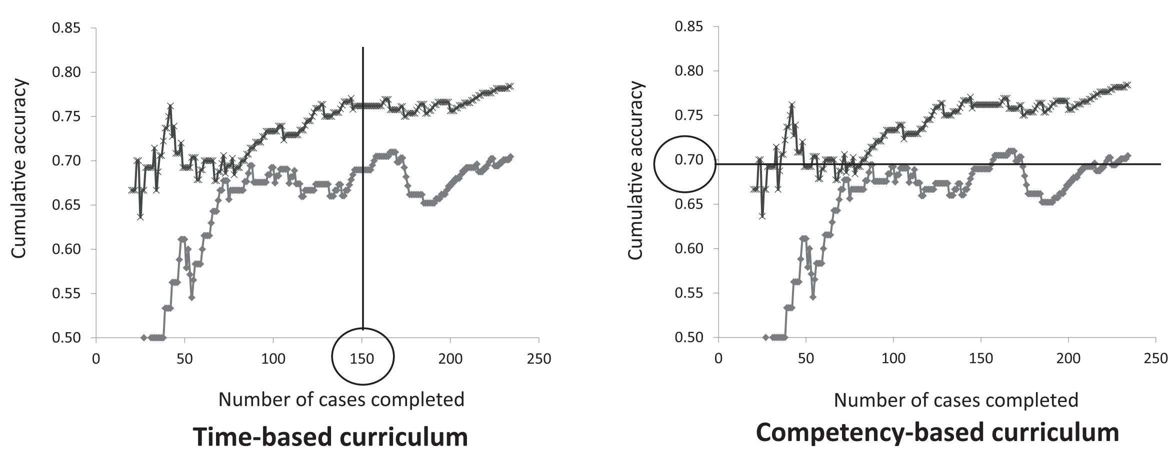

Figure 7 Traditional time-based learning compared with competency-based learning. Two residents are learning radiograph interpretation by practicing cases with feedback. Each resident improves with deliberate practice, but they start at different initial levels and progress at different rates, even though they are both at the same training level. In a time-based curriculum (left panel), time/effort is fixed while terminal competency is allowed to vary, and so, if each is assigned 150 cases, they finish with considerably different levels of competence. In a competency-based curriculum (right panel), each resident spends the amount of time required to achieve a given level of competence so that while the time/effort necessary to achieve competence varies considerably, the terminal competency standard is consistent.

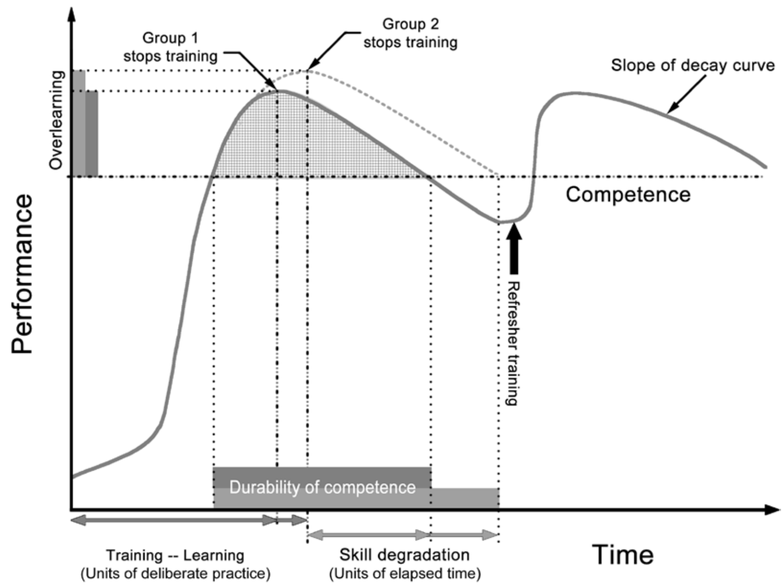

Figure 8 Prototypical experience curve. An experience curve incorporates both a learning curve, which describes the increase in performance during a training period, and a forgetting or skill degradation curve, which describes the decrease in skill that occurs when it is not practiced. This model can help guide refresher training schedules. Adapted with permission from Pusic M, Kessler D, Szyld D, Kalet A, Pecaric M, Boutis K. Experience curves as an organizing framework for deliberate practice in emergency medicine learning. Acad Emerg Med. 2012;19:1476–1480.

Potential Pitfalls and Limitations

There are several potential pitfalls to consider when using learning curves. First, learning curve models, like all statistical models, are approximations susceptible to errors and confounding. Model specification error will arise when the incorrect learning curve formula is chosen for a given situation, although testing for goodness of fit can mitigate the possibility of selecting the wrong formula.[33] Another bias, omitted-variable bias, occurs when important aspects represented in the construct of "effort" are not represented by the learning curve. Important examples of potentially omitted variables include level of learner motivation, perceived value of the task, degree of self-regulation, and emotional state.[34] Further, learning environment factors, such as situational fidelity, scheduling, and rewards systems, also influence learning.[35] Educators can optimize accuracy and validity by collecting as much information as is practical and incorporating it into multivariable learning curve models. Examining multiple variables is greatly facilitated by so-called "big data" approaches to collecting and analyzing educational data.[36] However, learning curves still remain especially susceptible to confounding because confounding can occur, as discussed, at the level of the measure of effort, the measure of performance, or with the linking function.

Another type of confounding occurs if the instructional method interacts with the measure of effort. For example, fatigue or boredom can increase with each repetition, which can result in an apparent flattening of the learning curve compared with what it would look like under optimal circumstances. Floor and ceiling effects in which assessment questions are either, respectively, too easy or too hard result in learning curves that are artificially flat.

Although we have used a validity framework to suggest how to generate a meaningful learning curve, the utility of applying valid learning curves within a given educational setting is a separate question that requires further research. Thus far, few researchers have examined how the specific shape of any given learning curve can serve to identify problem learners. If hundreds of repetitions of a given learning task are required to generate a reliable and valid learning curve, then educators whose curricular programs are already filled to bursting will have to be selective as to when they should use learning curve models. Also, in view of the susceptibility to confounding described above, the applicability of learning curves to complex learning situations will likely be more difficult. Our example learning intervention, radiograph interpretation, is useful because each case represents a defined unit of effort, a large number of cases are available, each case results in a relatively dichotomous answer, feedback is easy to deliver and measure, the performance can be reliably measured, and the administration of cases over time can be consistent across learners. These ideal circumstances will not apply to a majority of the to-be-learned clinical competencies. Our contention is that these limits need to be explored, especially now, in a time when better efficiency of data collection may allow big data approaches to some of the logistics of generating meaningful learning curves.[36]

Final Thoughts: Learning Curves as a Metaphor

As useful as they might be as instructional devices or assessment metrics, learning curves also operate at a general level as an organizing metaphor for knowledge and skill acquisition. First, the curves lay bare individual variability in learners' attainment of an educational goal through standardized instructional designs. Different students who receive the same instruction clearly experience the instruction differently and construct their knowledge according to their own unique set of experiences and ways of seeing the world.[37] Second, the learning curve representation of effort leading inexorably to goal achievement applies to many health education settings. Additionally, an assessment paradigm that explicitly includes effort sends a positive educational message that is consistent with the growth of mind-sets correlated with resilience and ultimate achievement.[38] Although the starting point and slope might vary, a learning curve's message is ultimately a hopeful one: namely, the paths of those who came before can inform present learners and attest that, though each path is unique, hard work can reliably enable outstanding levels of individual achievement. By reinforcing this message with concrete examples of learning curves, we can see the added benefit of aligning our educational culture with these values.

Acknowledgments: The authors wish to thank the members of the Research on Medical Education Outcomes unit of the Program for Medical Education Innovation and Research at New York University Langone School of Medicine for their helpful reviews of this manuscript.

Funding/Support: The radiology studies[1,17] described in this paper were carried out with the support of Medical Education Research Grants from the Royal College of Physicians and Surgeons of Canada.

Other disclosures: None reported.

Ethical approval: The authors of all the original studies[1,17] described obtained ethical approval to conduct their work.

Previous presentations: Some of this content was previously presented at the Association of American Medical Colleges Research in Medical Education Annual Meeting, November 2009, Boston, Massachusetts, and at the Pediatric Academic Societies Annual Meeting, May 2009, Baltimore, Maryland.

M.V. Pusic is assistant professor, Emergency Medicine, and director, Division of Education Quality and Analytics, New York University Langone School of Medicine, New York, New York.

K. Boutis is associate professor and pediatric emergency physician, Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

R. Hatala is associate professor, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

D.A. Cook is professor, Medicine and Medical Education; director, Online Learning Development and Analysis, Center for Online Learning, Mayo Clinic College of Medicine; and consultant, Division of General Internal Medicine, Mayo Clinic, Rochester, Minnesota.

References

1 Pusic M, Pecaric M, Boutis K. How much practice is enough? Using learning curves to assess the deliberate practice of radiograph interpretation. Acad Med. 2011;86:731–736.
2 Ericsson KA. Deliberate practice and acquisition of expert performance: A general overview. Acad Emerg Med. 2008;15:988–994.
3 Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. Health Technol Assess. 2001;5:1–79.
4 Ritter FE, Schooler LJ. The learning curve. In: Smelser NJ, Baltes PB, eds. International Encyclopedia of the Social and Behavioural Sciences. 1st ed. Oxford, UK: Elsevier Science Ltd.; 2002:8602–8605.
5 Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—Rationale and benefits. N Engl J Med. 2012;366:1051–1056.
6 McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the evidence. Acad Med. 2011;86:706–711.
7 Cook DA, Andriole DA, Durning SJ, Roberts NK, Triola MM. Longitudinal research databases in medical education: Facilitating the study of educational outcomes over time and across institutions. Acad Med. 2010;85:1340–1346.
8 Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73:13–22.
9 Son LK, Sethi R. Adaptive learning and the allocation of time. Adapt Behav. 2010;18:132–140.
10 Singer JD, Willett JB, eds. Longitudinal Data Analysis: Modeling Change and Event Occurrence. 1st ed. New York, NY: Oxford University Press; 2003.
11 Kane MT. Current concerns in validity theory. J Educ Meas. 2001;38:319–342.
12 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. Am J Med. 2006;119:166.e7–166.e16.
13 Downing SM. Validity: On meaningful interpretation of assessment data. Med Educ. 2003;37:830–837.
14 Van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach. 1996;18:103–110.
15 American Educational Research Association. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association; 1999.
16 Dinçler S, Koller MT, Steurer J, Bachmann LM, Christen D, Buchmann P. Multidimensional analysis of learning curves in laparoscopic sigmoid resection: Eight-year results. Dis Colon Rectum. 2003;46:1371–1378.
17 Boutis K, Pecaric M, Seeto B, Pusic M. Using signal detection theory to model changes in serial learning of radiological image interpretation. Adv Health Sci Educ Theory Pract. 2010;15:647–658.
18 Altman DG. Practical Statistics for Medical Research. London, UK: Chapman & Hall/CRC; 1991.
19 Bienkowski M, Feng M, Means B. Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. Washington, DC: U.S. Department of Education, Office of Technology; October 2012. http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf. Accessed January 2, 2015.
20 Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. Stat Methods Med Res. 2003;12:147–170.
21 Holzhey DM, Jacobs S, Walther T, Mochalski M, Mohr FW, Falk V. Cumulative sum failure analysis for eight surgeons performing minimally invasive direct coronary artery bypass. J Thorac Cardiovasc Surg. 2007;134:663–669.
22 Schuwirth LW, Van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33:478–485.
23 Regehr G. It's NOT rocket science: Rethinking our metaphors for research in health professions education. Med Educ. 2010;44:31–39.
24 Dreyfus SE, Dreyfus HK. A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Berkeley, Calif: University of California, Berkeley; 1980.
25 Ericsson K. The acquisition of expert performance as problem solving. In: Davidson J, Sternberg R, eds. The Psychology of Problem Solving. 1st ed. Cambridge, England: Cambridge University Press; 2003:44–83.
26 Green ML, Aagaard EM, Caverzagie KJ, et al. Charting the road to competence: Developmental milestones for internal medicine residency training. J Grad Med Educ. 2009;1:5–20.
27 Hodges B. The shifting discourses of competence. In: Hodges BD, Lingard L, eds. The Question of Competence: Reconsidering Medical Education in the Twenty-First Century. London, UK: Cornell University Press; 2012:1–32.
28 Brydges R, Butler D. A reflective analysis of medical education research on self-regulation in learning and practice. Med Educ. 2012;46:71–79.
29 Quirk M. Metacognitive capabilities. In: Quirk ME, ed. Intuition and Metacognition in Medical Education. 1st ed. New York, NY: Springer; 2006:23–35.
30 Kornell N, Metcalfe J. Study efficacy and the region of proximal learning framework. J Exp Psychol Learn Mem Cogn. 2006;32:609–622.
31 Guadagnoli M, Morin MP, Dubrowski A. The application of the challenge point framework in medical education. Med Educ. 2012;46:447–453.
32 Pusic MV, Kessler D, Szyld D, Kalet A, Pecaric M, Boutis K. Experience curves as an organizing framework for deliberate practice in emergency medicine learning. Acad Emerg Med. 2012;19:1476–1480.
33 Motulsky H, Christopoulos A. Approach to comparing models. In: Motulsky H, Christopoulos A, eds. Fitting Models to Biological Data Using Linear and Nonlinear Regression. 1st ed. Oxford, UK: Oxford University Press; 2004:134–151.
34 Song HS, Kalet AL, Plass JL. Assessing medical students' self-regulation as aptitude in computer-based learning. Adv Health Sci Educ Theory Pract. 2011;16:97–107.
35 Ellaway RH, Pusic M, Yavner S, Kalet AL. Context matters: Emergent variability in an effectiveness trial of online teaching modules. Med Educ. 2014;48:386–396.
36 Ellaway RH, Pusic MV, Cameron T, Galbraith R, Farrell J. Developing the role of big data and analytics in medical education. Med Teach. 2014;36:216–222.
37 Jonassen D. Objectivism versus constructivism: Do we need a new philosophical paradigm? Educ Technol Res Dev. 1991;39:5–14.
38 Dweck CS, Chiu C, Hong Y. Implicit theories and their role in judgments and reactions: A word from two perspectives. Psychol Inq. 1995;6:267–285.