Contents lists available at ScienceDirect

Surgery

journal homepage: http://ees.elsevier.com

A machine learning approach to predict surgical learning curves

Yuanyuan Gao Meng[a], Uwe Kruger EngD[a,b], Xavier Intes PhD[a,b], Steven Schwaitzberg MD[c,d,e], Suvranu De ScD[a,b,*]

[a] Center for Modeling, Simulation and Imaging in Medicine, Rensselaer Polytechnic Institute, Troy, NY
[b] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY
[c] Jacobs School of Medicine and Biomedical Sciences, The State University of New York, Buffalo, NY
[d] Department of Surgery, The State University of New York, Buffalo, NY
[e] Buffalo General Hospital, NY

ARTICLE INFO

Article history:
Accepted 21 October 2019
Available online xxx

ABSTRACT

Background: Contemporary surgical training programs rely on the repetition of selected surgical motor tasks. Such methodology is inherently open ended with no control on the time taken to attain a set level of proficiency, given the trainees' intrinsic differences in initial skill levels and learning abilities. Hence, an efficient training program should aim at tailoring the surgical training protocols to each trainee. In this regard, a predictive model using information from the initial learning stage to predict learning curve characteristics should facilitate the whole surgical training process.

Methods: This paper analyzes learning curve data to train a multivariate supervised machine learning model. One factor is extracted to define the trainees' learning ability. An unsupervised machine learning model is also utilized for trainee classification. When established, the model can predict robustly the learning curve characteristics based on the first few trials.

Results: We show that the information present in the first 10 trials of surgical tasks can be utilized to predict the number of trials required to achieve proficiency ($R^2 = 0.72$) and the final performance level ($R^2 = 0.89$). Furthermore, only a single factor, learning index, is required to describe the learning process and to classify learners with unique learning characteristics.

Conclusion: Using machine learning models, we show, for the first time, that the first few trials contain sufficient information to predict learning curve characteristics and that a single factor can capture the complex learning behavior. Using such models holds the potential for personalization of training regimens, leading to greater efficiency and lower costs.

Introduction

Bimanual motor skill learning is an important aspect of surgical training. Surgeons learn technical skills through repeated practice. However, most residency programs provide the opportunity to practice skills without ensuring that a certain level of proficiency has been reached. Technical surgical skills have been traditionally assessed using in-training evaluation reports, procedural-based assessment, or surgical logbooks. All these approaches are based on the traditional model of apprenticeship with the faculty responsible for assessing technical proficiency based on direct observation. However, problems including leniency/severity errors, central tendency errors, and "halo effects" associated with such approaches are well known.[1,2] Moreover, trainees have inherent differences in initial skill levels and learning rates and variations between surgical procedures, which limits the effectiveness of the time-limited training approach.

Technical skill testing for certification and competency-based curricula are increasingly popular. Demonstrating proficiency in basic laparoscopic and endoscopic skills is now a prerequisite for certification in general surgery.[3] Starting in 2018, the fundamentals of laparoscopic surgery (FLS) program is also required for board certification for obstetric and gynecological surgery (www.flsprogram.org). Realizing the inherent problems with the approach of repeated practice, there is significant interest in proficiency-based training.[4–7] In this approach, repetition is continued until a certain level of proficiency is achieved. However, the procedure is cumbersome and time consuming, as the number of repetitions is not known in advance. To develop structured training programs that account for individual variability in skills and learning abilities, a more personalized method is needed, which can predict individual learning curves for any surgical procedure based on initial performance. This requires a deeper understanding of learning curves for surgical skills.

[*] Reprint requests: Suvranu De, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180.
E-mail address: des@rpi.edu (S. De)

https://doi.org/10.1016/j.surg.2019.10.008
0039-6060/© 2019.

Different techniques to analyze and model the learning curves of surgical procedures have been presented in the literature. A review paper[8] has summarized those approaches. First, without any statistical analysis, a simple graph or a table displaying the outcome of the surgery against the number of operations was presented to show the learning curve in a substantial number of studies.[8] As a higher level of analysis, basic statistical tools such as *t*-test, analysis of variance, or χ^2 test was applied to 2 or 3 groups of data split by a number of practices.[8] However, the splitting points in these studies were arbitrarily selected, and the underlying curves in each group were not analyzed.[8]

Besides statistically analyzing the learning curves, analytical modeling methods have also been presented in literature. A commonly used modeling method is to fit a curve to the learning curve data using least squares regression, with or without an adjustment for other confounding factors including age and sex.[8] Both linear and exponential curves have been used to fit learning curves, without much justification for the choice of these curves.[8]

The cumulative summation (CUSUM) technique was initially suggested to monitor surgical performance[9] but was recently applied to analyze the learning curve of a surgical skill.[10–15] CUSUM has a simple formulation in that positive or negative increments are added to a cumulative score according to failure or success of the successive trial.[16] A graph representation of CUSUM is intuitive in that a declining trend indicates successive successes and an increasing trend indicates successive failures.[16] The 2 boundary limits, h_1 and h_0.in CUSUM graphs represent whether or not the observed failure rate is significantly different from the desired acceptable failure rate.[16] The number of trials to achieve proficiency is derived by counting the number of attempts before crossing the boundary limits.[17] With different specific aims, the design of CUSUM schemes varies across studies. In some studies,[18–20] CUSUM is calculated as the cumulative difference between observed and expected outcome values, such as operation time or blood loss, instead of binary results of success or failure. A change-point in these CUSUM graphs was determined to derive the number of trials to achieve proficiency[18,20] or learning phases.[19] Some studies set the downward part of CUSUM graphs to 0 to monitor only failures.[21]

A unique disadvantage of all these existing approaches is that they require information regarding training that has already taken place. Although they are academically of interest, their utility in designing individualized training programs is limited. To the best of our knowledge, there are no approaches in the literature that focus on predicting learning curve features, including the final performance level after a certain number of trials. Hence, our goal in this study is to overcome these limitations in the existing literature by testing 2 hypotheses. First, we hypothesize that the performance of a trainee during the first few trials of a bimanual motor task has sufficient information to predict the number of trials required to achieve proficiency and the final performance level. Second, we hypothesize that it is possible to define a single factor that can describe how these parameters, including the initial skill level of the trainee, are related to each other. To accomplish that, we have performed a meta-analysis of bimanual skill acquisition in a pattern cutting task, which is a part of the FLS program.[13,15] Learning data on the physical FLS trainer box and on a virtual basic laparoscopic skill trainer (VBLaST) replicating the FLS tasks[14,22–25] have been utilized.

Materials and methods

Data sources

In this study, we performed a retrospective analysis of the learning curve data from 3 IRB approved studies.[13,15,26]. The studies were associated with the FLS pattern cutting task which involves cutting a gauze following a marked circle[13,15]and the FLS intracorporeal suturing.[26] The 3 studies were selected because of the similarity in experimental setup and procedures. Novice medical students, with no prior surgical experience, were selected for these studies.[13,15,26] The training was carried out over a 3-week period using the FLS trainer box or VBLaST replicating these tasks. The FLS program has been shown to be reliable to quantify surgical skill level,[27,28] and metrics for these tasks are established in the surgery literature[29] and used in board certification in general surgery. The performance score of each trial was calculated from task performance time and performance error using the accredited FLS scoring methodology with consent under a nondisclosure agreement from the FLS Committee. The metrics for the VBLaST are derived from the FLS metrics and discussed elsewhere.[23] In both cases, the metrics are aggregated into a final score, either being the FLS score or the VBLaST score. These 2 scores are the quantities used to build the surgical learning curves. We excluded as outliers those curves that did not exhibit a clear initial learning stage or a learning plateau. Fifteen learning curves were then selected for this study (Table I).

Variables

This study involved 3 variables: the number of trials required to achieve proficiency, the initial performance level, and the final performance level. We defined the number of trials required to achieve proficiency based on the Technical Skill Proficiency-Based Training Curriculum[4] for the FLS program. For the pattern cutting task, proficiency is achieved if the task can be performed within 98 seconds on 2 consecutive repetitions; for the intracorporeal suture, trainees achieve proficiency when they perform 10 additional trials after 2 consecutive trials within 112 seconds with allowable errors. A similar definition was employed with the VBLaST data. We defined the initial performance level as the average score of the first 3 trials. To define the final performance level, multiple 2-tailed *t*-tests were performed between different trial intervals. From the result in Table II, the fifth 10 trials performance is not significantly different from the fourth 10 trials, but the other pairs of trial intervals are significantly different. This indicates that from the 40th trial, the performance scores do not significantly change. Thus, we defined the final performance level as the average score after the 40th trial.

Multivariate supervised learning model

To test our first hypothesis, we used a multivariate supervised machine learning approach known as kernel partial least squares (KPLS)[30,31] to predict the learning curve features, including the number of trials to achieve proficiency and the final performance level

Table I
The source of the data [13, 15]

Study	Platform	Motor task	No. of learning curves
Nemani et al 2017 [13]	FLS *	Pattern cutting	4
Nemani et al 2017 [13]	VBLaST *	Pattern cutting	6
Linsk et al 2017 [15]	VBLaST *	Pattern cutting	2
Fu et al 2019 [26]	FLS *	Suturing	3

* FLS is a physical training box and VBLaST is a virtual reality version of it.

Table II
Significant test results between trial intervals

Trial intervals	Significant test
First vs second 10 trials	$P = .000$ *
Second vs third 10 trials	$P = .004$ *
Third vs fourth 10 trials	$P = .023$ *
Fourth vs fifth 10 trials	$P = .205$

* Significance $\alpha = 0.05$.

from the initial learning performance. KPLS first computes a non-linear transformation of the 10 initial trials' performance scores (denoted as X) from the 13 learning curve entries into a high dimensional feature space. Once transformed, the projected observations are then linearly regressed to maximize their covariance with the final performance score or the number of trials to achieve proficiency (denoted as y). In this way, the non-linear relationship could be modeled optimally between the input X and the output y.

We used the coefficient of determination (denoted as R^2) as defined below to quantify the accuracy of the proposed model:

$$R^2 = 1 - \frac{SSE}{SST} \tag{2}$$

$$SSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-1} \tag{3}$$

$$SST = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} \tag{4}$$

where y_i is the true outcome value, $\hat{y}_i$ is the predicted outcome value, $\bar{y}$ is the mean of the true outcome value, and n is the total number of samples. $R^2 = 1$ indicates that all the variance of the data is explained by the model.[30] Conversely, smaller values or even negative values indicate a poor predictability is considered.

Considering the small sample size of the dataset, we validated our modeling results using the leave-one-out cross-validation scheme.[31] In leave-one-out cross-validation, we excluded 1 learning curve from the training dataset and tested the model on the learning curve left out, repeating the analysis for all learning curves to ensure robustness of the model and to avoid overfitting.

Because the dataset is from subjects practicing on 2 types of training platforms, a physical FLS trainer box, and a VBLaST, it is necessary to validate whether the model works across different platforms. To achieve this, a platform-wise, cross-validation scheme was designed. In this scheme, we trained the model on one platform and tested it on the other one to demonstrate the effectiveness of the model across the 2 platforms.

Wright[32] was the first to discuss learning curve modeling, and his log-linear model has been used in prior literature.[33] Other log-linear based models have also been developed considering different factors affecting learning curves, like the S-Curve model which takes a gradual start-up into consideration.[33] In the experimental setup in this study, the subjects recruited were novice medical school students with no prior surgical experience, and the training paradigm was consistent across trials. Thus, we adopted the conventional log-linear model as a comparison, given by the following equation:

$$Y = Y_0 N^{\theta} \tag{5}$$

where N is the number of trials, Y is the performance, θ is the learning rate, and Y_0 is the initial performance level. The model variables are independently assessed for each subject.

Factor analysis model

Factor analysis is a statistical method to explore latent variables or factors from the observed variables. In our case, we have 3 observed variables—initial performance level, number of trials to achieve proficiency, and final performance level. The question we asked is whether a single variable can be used to represent all 3 variables. Here, we use a factor analysis model known as the kernel principal component analysis (KPCA) approach[31] to extract a representative factor, learning index (LI), from the 3 learning curve features. We first mapped the 3 features into a high dimensional non-linear space. Then, we extracted 1 principal component as LI from these high-dimensional data by principle component analysis. To test whether LI represents the 3 learning curve features, we used LI to predict the 3 features by KPLS. If LI could predict all 3 features accurately, then the information compressed in LI is enough to represent the 3 learning curve features.

Unsupervised classification of skill level

An unsupervised learning approach known as k-means clustering analysis[34,35] was adopted to separate the trainees according to their different learning curve characteristics. By analyzing the learning curve features of different trainee groups, the learning characteristics of each group were summarized. Furthermore, grouping results derived from LI was compared to the grouping result from all the 3 features to see whether LI could indicate unique learning characteristics.

Results

Learning curve data and features

The learning curve data for all subjects are presented (4 trained on FLS physical training box and 9 trained on VBLaST virtual reality trainer) in Fig 1 and the 3 features calculated from the learning curves are summarized in Table III.

Prediction performance

Table IV represents the R^2 values for KPLS and the log-linear models when they are used to predict the number of trials to achieve proficiency and the final performance level based on the performance in initial several trials. The performance of the KPLS model indicates that the initial performance pattern can be used to predict the 2 learning curve features with a high degree of accuracy. Conversely, using the log-linear model, the resultant R^2 values for predicting the number of trials to achieve proficiency are negative and those for predicting the fi-

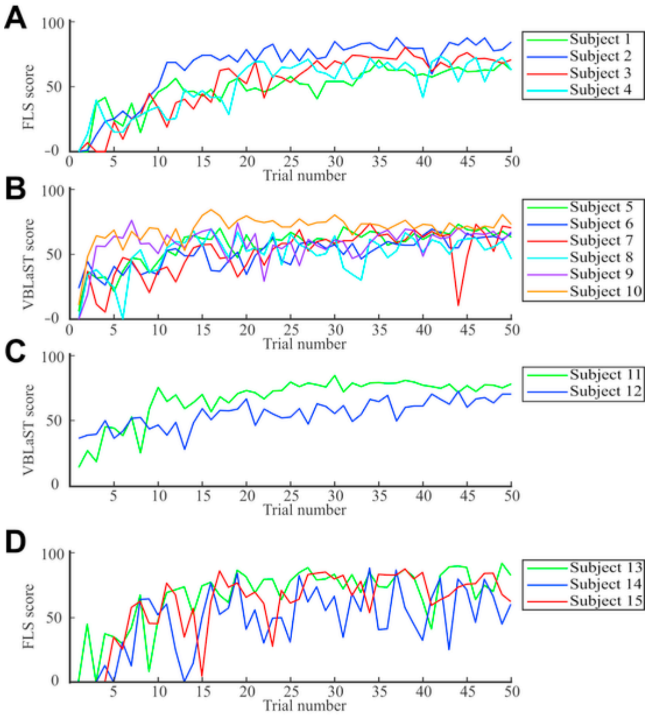

Fig 1. The learning curves are from 3 studies: (*A*) FLS pattern cutting study in Nemani et al 2017[13]; (*B*) VBLaST pattern cutting study in Nemani et al 2017[13]; (*C*) VBLaST pattern cutting study in Linsk et al 2017[15]; and (*D*) FLS intracorporeal suture study in Fu et al 2019.[26] (The permission to reuse the data has been acquired from the journals).

Table III
The learning curve features values for all subjects (performance levels refer to the FLS/VBLaST scores)

Subject no.	Initial performance level (average from trial #1 to trial #3)	Number of trials required to achieve proficiency	Final performance level (average from trial #40 to trial #50)
1	13.10	57	65.46
2	4.88	15	83.36
3	2.38	33	72.86
4	18.57	22	69.32
5	26.07	45	70.50
6	34.88	114	66.84
7	16.90	50	62.28
8	25.36	67	59.64
9	26.31	52	68.21
10	42.98	10	74.93
11	21.43	20	80.32
12	40.83	50	71.14
13	16.13	56	83.50
14	0	114	66.77
15	0	97	76.95

Table IV
Accuracy of KPLS and log-linear model

Learning curve variable	KPLS	Log-linear model
Number of trials to achieve proficiency	$R^2 = 0.72$ (first 10 trials)	$R^2 = -4.21$ (first 50 trials) * $R^2 = -\ 9.27$ (first 40 trials) * $R^2 = -15.17$ (first 30 trials) * $R^2 = -49.53$(first 20 trials) * $R^2 = -109.55$ (first 10 trials) *
Final performance level	$R^2 = 0.89$ (first 10 trials)	$R^2 = 0.76$ (first 50 trials) $R^2 = 0.58$ (first 40 trials) $R^2 = 0.37$ (first 30 trials) $R^2 = -0.27$ (first 20 trials) $R^2 = -3.36$ (first 10 trials)

* For the first 10 trials, the predicted learning curves of subject 4, 6, 12, and 13 do not achieve proficiency in 1,000 trials and are excluded from the

R^2

R^2
calculation; for the first 20 trials, the predicted learning curve of subject 6 does not achieve proficiency in 1,000 trials and is excluded from the

R^2

R^2
value calculation; for the first 30, 40, and 50 trials, the predicted learning curve of subject 14 does not achieve proficiency in 1,000 trials and is excluded from the

R^2

R^2
value calculation.

nal performance level indicate the need for a considerable number of initial trials to be close to 1.

The log-linear model performance was further explored with a different number of initial trials for which data were used for model development. One example of the learning curve from subject 1 is shown in Fig 2. The figure shows that the log-linear model works well when the FLS scores for the first 50 trials are known. However, when fewer trials are used as input to the model, the log-linear curve clearly becomes less accurate. This is particularly evident when data from only the first 10 trials are used, and the log-linear curve underestimates the learning effect considerably. The underestimation of the log-linear model also explains why the R^2 values for predicting the number of trials to achieve proficiency are all negative using this approach (Table IV). The predicted curves are grossly underestimated, implying that a much larger number of trials is required to achieve proficiency than what is actually needed. Some predicted curves could not achieve proficiency even after 1,000 trials and are excluded from the analysis (Table IV). Although the variables being included in this work are very simple, it is shown to be hard to predict those using existing models, such as the log-linear model.

Considering the 2 different platforms the subjects practiced on (physical and virtual), platform-wise, cross-validation tests were performed on the KPLS model. The results are listed in Table V. When predicting the number of trials to achieve proficiency and the final performance, the R^2 of the KPLS model are all above 0.70 in the platform-wise, cross-validation testing. This indicates that the learning curve data patterns are consistent across the physical FLS training box and the VBLaST box, and the datasets could be meta-analyzed.

Factor analysis

We derived a single factor, which we referred to as LI, from the 3 learning curve features (initial performance level, number of trials to achieve proficiency, and final performance level), based on the KPCA method. LI is a latent variable which depicts the learning characteristics of the learners. The R^2 values when using LI to predict the 3 learning curve features by a KPLS model are listed in Table VI. Since all the R^2 values are above 0.8, the extracted feature could be determined as a representative feature of the 3 features.

K-means clustering

Next, we further grouped the subjects by their learning curve features (initial performance level, number of trials to achieve proficiency, and final performance level) using the k-means clustering algorithm. Two groups naturally emerged from this analysis: group 1 with subjects 2, 10, 11, and 13 and group 2 consisting of the remaining ones. To understand the implication of this grouping, we plotted the number of trials to achieve proficiency and the final performance level against the initial performance level in Fig 3, *A*. The crosses represent subjects in group 1, and the circles represent subjects in group 2. From these plots, it is clear that trainees in group 1 have higher initial performance levels, require fewer trials to achieve proficiency, and achieve higher final performance levels; whereas the trainees in group 2 have lower initial performance levels, need more trials to achieve proficiency, and achieve lower final performance levels. When we use the same k-means clustering algorithm to group the subjects based on the feature values of LI, the same grouping result is derived as shown in Figure 3, *B*. The 2 groups are clearly separable and clustered by the extracted features. This result supports that LI is sufficient to classify learners with unique learning characteristics.

Discussion

Our findings highlight that the use of machine learning enables assessing the performance of a trainee by evaluating his or her performance during the first 10 repetitive trials. Based on only 10 trials, we can predict (1) the number of trials required to achieve proficiency and (2) the final performance level, as defined as the average FLS score after 40th trial, with a high degree of accuracy. Furthermore, a single factor, LI, which we refer to as the learning ability, can be derived from a non-linear factor analysis model. The single factor describes common variation within these 2 parameters and the initial performance level. This, in turn, implies that the number of trials required to achieve pro-

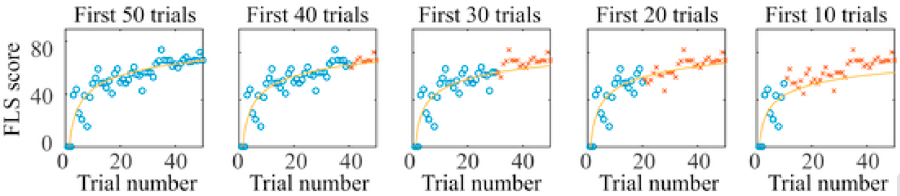

Fig 2. The performance of the log-linear model developed using the scores for the first 50, 40, 30, 20, and 10 of trials for subject 1. "o"s represent the trials for which the scores are assumed to be known; "x"s represent the remainder of the trials for which the data are not used in developing the model. The solid line represents the log-linear model.

Table V
Cross-validation results of KPLS

Learning curve variable	Platform-wise	
	Trained on FLS platform but tested on the VBLaST platform	Trained on the VBLaST platform but tested on FLS platform
Number of trials to achieve proficiency	$R^2 = 0.72$	$R^2 = 0.78$
Final performance level	$R^2 = 0.73$	$R^2 = 0.78$

Table VI
The R^2 values using the learning ability to predict the 3 learning curve features

Case	R^2
Predicting all the 3 features	0.92
Initial performance level	0.87
Number of trials to achieve proficiency	0.93
Final performance level	0.94

ficiency, the final performance level, and the initial performance are not independent of each other.

These findings are related to earlier work in this area, as the initial learning stages are related to the later stages.[36,37] Moreover, Jirapinyo et al[38] showed that a log-linear regression model can describe surgical training learning curves reasonably well when all learning data are used. However, other studies contradict this.[19,39] With respect to our study, even though we extracted very simple learning curve characteristics, we still found limitation in the use of log-linear models. One potential reason is that the log-linear regression model is derived from group data, but individual learning progress may be distinct from the group trend.[40] Another, more intuitive reason is that the log-linear model assumes a predefined form for the learning curves before fitting the curves. In sharp contrast, the KPLS model does not make such an assumption. Additionally, it is well known that a larger set of variables that are highly correlated or collinear can yield difficulties in identifying regression models. KPLS is a nonlinear regression tool that has been developed in the field of chemometrics with the aim to handle such situations.[30] Therefore, the KPLS regression model is able to capture the complex process of surgical skill learning in a data-driven format, which is missed by simplistic analytical models including the log-linear model. Once a KPLS regression model has been developed for a surgical task, the trained model may, consequently, be employed to predict the number of trials needed to achieve proficiency by any new learner based on scores of the initial few trials.

Moreover, the use of machine learning for learning curve prediction has the potential to redesign surgical training programs and has implications for skill decay and retraining throughout the entire professional life of surgeons. Predicting the learning curve variables early in the training process would help to provide more focused feedback and implement adaptive learning strategies. The idea of adaptive training is not new and has been suggested in the surgical literature. For example, Stefanidis et al[41] pointed out that by establishing skill learning curves, the training curricula could be tailored to provide additional training to those who need more training than others. Another study[42] compared adaptive curricula and volume-based curricula in surgical training and demonstrated that the group trained with adaptive curricula achieved the same level of performance but required fewer training hours.

Another important finding of our study is that the single factor allows clustering the trainees into 2 groups based on their distinct learning curve characteristics. These groups have clearly unique characteristics, where the participants in group 1 had a higher initial ability to carry out the FLS task, showed higher final performance level, and did not require a considerable number of trials to converge from the ini-

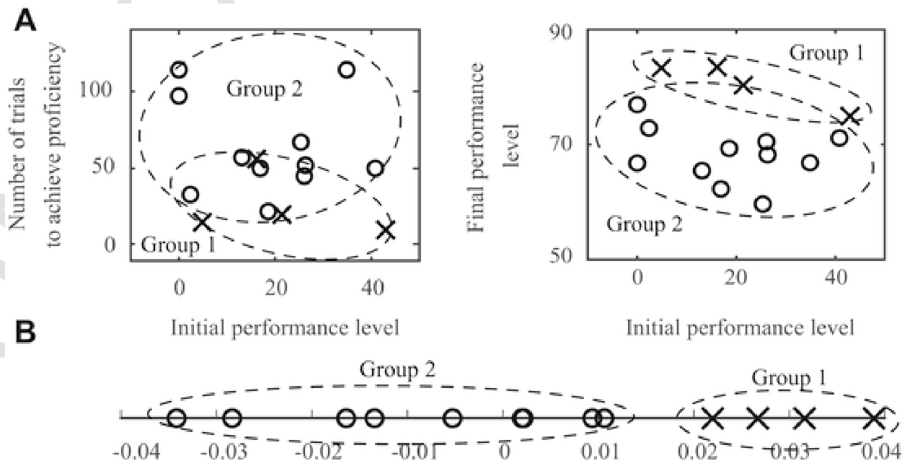

Fig 3. The trainees were clustered into 2 groups by the k-means clustering algorithm. "x"s represent trainees that are clustered into group 1 and "o"s represent trainees that are clustered into group 2. The grouping results are from (*A*) the learning curve features and (*B*) the extracted factor LI.

tial to their final performance level. On the other hand, the subjects in group 2 had a lower initial ability, showed a lower final performance level, and required more repetitive trials to reach the final performance level. The surgical motor learning difference between individuals has also been demonstrated and studied in other studies. For example, Louridas et al[43] showed that with the same training curriculum, the participants demonstrated different learning results and could be divided into 3 groups with top, moderate, and low performance. Similar grouping results of surgical residents were reported.[44] Our study provides a quantitative understanding that individuals with different initial skill levels require different practice to reach a final performance level. Differences in learning characteristics may be due to innate factors including handedness, gender, visual-spatial ability, and confidence level[45–51] or extrinsic factors including research experience, selection of specialty, and grouping.[49,52,53]

Our study has several limitations which may suggest conducting further research in this area. First, the learning curves were quantified by task performance scores, which are calculated from performance time and performance error. Although these same metrics are used in the FLS, there could be other kinematic (eg, hand trajectory) or physiological (eg, eye motion or skin conductance) metrics which could provide further information regarding performance. In separate work, we have shown that functional brain imaging that relies on neuro-vascular coupling provides much more accurate quantification of bimanual motor skill learning than the traditional FLS metrics.[54,55] Second, the bimanual motor task in this study is limited to the pattern cutting task and intracorporeal suture task. Extending the analysis procedures to other FLS tasks and actual surgical procedures is left for future work. It is important to note that the results reported can assist in planning individualized training regimes. It is not intended to share the predicted number of trials with the learners, as this may affect their performance negatively. In addition to that, physiological measurement, such as functional brain imaging, could play an important role to monitor the workload and attention level to determine whether the trainees are trying their best to learn. Finally, data from only a few learning curves have been used in this study. Learning curve studies are inherently difficult to do owing to the extended time commitment of the medical students and the limited number of willing participants. As summarized in a comprehensive review paper,[56] a sample size of 8 to 23 is reported in previous simulation-based surgical training studies, indicating the difficulty to recruit participants in multiple days training protocols. We included 3 of our previous studies collecting the learning curve information to have a sample size of 15. To set up a machine learning model based on such a small sample size, we selected KPLS, which is a machine learning method suitable for small sample size. Such methods have an extensive record of applications in data chemometrics,[57–59] where small sample sizes are the norm rather than the exception. Moreover, we report results that are based on an independent assessment of the model performance to ensure that the established regression models are not compromised by overfitting. Future multi-center studies may mitigate some of these issues.

In conclusion, we propose the use of sophisticated machine learning models for predicting the learning curve features from the initial few trials of bimanual surgical motor tasks. A single factor, LI, which we define as the learning ability, can capture complexities of learning behavior. Use of such models holds the potential for personalization of training regimens, leading to greater efficiency and lower costs.

Funding/Support

We thank the funding provided by NIH / National Institute of Biomedical Imaging and Bioengineering grants 2R01EB005807, 5R01EB010037, 1R01EB009362, 1R01EB014305, and R01EB019443.

Conflict of interest/Disclosure

Yuanyuan Gao, Uwe Kruger, Xavier Intes, Steve Schwaitzberg and Suvranu De have no conflicts of interest or financial ties to disclose.

References

1. K Moorthy, Y Munz. Objective assessment of technical skills in surgery. BMJ. 2003;327:1032–1037.
2. J Shah, A Darzi. Surgical skills assessment: an ongoing debate. BJU Int. 2001;88:655–660.
3. N Soper, GM Fried. The fundamentals of laparoscopic surgery: its time has come. Bull Am Coll Surg. 2008;93:30–32.
4. DJ Scott, EM Ritter, ST Tesfay, EA Pimentel, A Nagji, GM Fried. Certification pass rate of 100% for fundamentals of laparoscopic surgery skills after proficiency-based training. Surg Endosc. 2008;22:1887–1893.
5. D Stefanidis, JR Korndorffer Jr, R Sierra, C Touchard, JB Dunne, DJ Scott. Skill retention following proficiency-based laparoscopic simulator training. Surgery. 2005;138:165–170.
6. G Ahlberg, L Enochsson, AG Gallagher, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg. 2007;193:797–804.
7. G Sroka, LS Feldman, MC Vassiliou, PA Kaneva, R Fayez, GM Fried. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room-a randomized controlled trial. Am J Surg. 2010;199:115–120.
8. CR Ramsay, AM Grant, SA Wallace, PH Garthwaite, AF Monk, IT Russell. Assessment of the learning curve in health technologies. a systematic review. Int J Technol Assess Health Care. 2000;16:1095–1108.
9. SH Steiner, RJ Cook, VT Farewell. Monitoring paired binary surgical outcomes using cumulative sum charts. Stat Med. 1999;18:69–86.
10. DJ Biau, M Resche-Rigon, G Godiris-Petit, RS Nizard, R Porcher. Quality control of surgical and interventional procedures: a review of the CUSUM. Qual Saf Health Care. 2007;16:203–207.
11. A Young, JP Miller, K Azarow. Establishing learning curves for surgical residents using Cumulative Summation (CUSUM) Analysis. Curr Surg. 2005;62:330–334.
12. S Bolsin, M Colson. The use of the Cusum technique in the assessment of trainee competence in new procedures. Int J Qual Health Care. 2000;12:433–438.
13. A Nemani, W Ahn, C Cooper, S Schwaitzberg, S De. Convergent validation and transfer of learning studies of a virtual reality-based pattern cutting simulator. Surg Endosc. 2018;32:1265–1272.
14. L Zhang, G Sankaranarayanan, VS Arikatla, et al. Characterizing the learning curve of the VBLaST-PT{©} (Virtual Basic Laparoscopic Skill Trainer). Surg Endosc. 2013;27:3603–3615.
15. AM Linsk, KR Monden, G Sankaranarayanan, et al. Validation of the VBLaST pattern cutting task: a learning curve study. Surg Endosc. 2018;32:1990–2002.
16. IG Kestin. A statistical approach to measuring the competence of anaesthetic trainees at practical procedures. Br J Anaesth. 1995;75:805–809.
17. DJ Biau, SM Williams, MM Schlup, RS Nizard, R Porcher. Quantitative and individualized assessment of the learning curve using LC-CUSUM. Br J Surg. 2008;95:925–929.
18. D Miskovic, M Ni, SM Wyles, P Tekkis, GB Hanna. Learning curve and case selection in laparoscopic colorectal surgery: systematic review and international multicenter analysis of 4852 cases. Dis Colon Rectum. 2012;55:1300–1310.
19. M Yoshida, N Kakushima, K Mori, et al. Learning curve and clinical outcome of gastric endoscopic submucosal dissection performed by trainee operators. Surg Endosc. 2017;31:3614–3622.
20. Z Wen, H Liang, J Liang, Q Liang, H Xia. Evaluation of the learning curve of laparoscopic choledochal cyst excision and Roux-en-Y hepaticojejunostomy in children: CUSUM analysis of a single surgeon's experience. Surg Endosc. 2017;31:778–787.
21. TO Lim, A Soraya, LM Ding, Z Morad. Assessing doctor's competence: application of CUSUM technique in monitoring doctors' performance. Int J Qual Health Care. 2002;14:251–258.
22. GP Moustris, SC Hiridis, KM Deliparaschos, KM Konstantinidis. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. Int J Med Robot. 2011;7:375–392.
23. A Chellali, W Ahn, G Sankaranarayanan, et al. Preliminary evaluation of the pattern cutting and the ligating loop virtual laparoscopic trainers. Surg Endosc. 2015;29:815–821.
24. G Sankaranarayanan, H Lin, VS Arikatla, et al. Preliminary face and construct validation study of a virtual basic laparoscopic skill trainer. J Laparoendosc Adv Surg Tech A. 2010;20:153–157.
25. VS Arikatla, G Sankaranarayanan, W Ahn, et al. Face and construct validation of a virtual peg transfer simulator. Surg Endosc. 2013;27:1721–1729.
26. Y Fu, L Cavuoto, D Qi, et al. Characterizing the learning curve of a virtual intracorporeal suturing simulator VBLaST-SS©. Surg Endosc. 2019 Available from:https://doi.org/10.1007/s00464-019-07081-6.
27. B Zendejas, RK Ruparel, DA Cook. Validity evidence for the Fundamentals of Laparoscopic Surgery (FLS) program as an assessment tool: a systematic review. Surg Endosc. 2016;30:512–520.
28. B Zendejas, JW Jakub, AM Terando, et al. Laparoscopic skill assessment of practicing surgeons prior to enrollment in a surgical trial of a new laparoscopic procedure. Surg Endosc. 2017;31:3313–3319.

29. SA Fraser, DR Klassen, LS Feldman, GA Ghitulescu, D Stanbridge, GM Fried. Evaluating laparoscopic skills, setting the pass/fail score for the MISTELS system. Surg Endosc. 2003;17:964–967.
30. K Kim, JM Lee, IB Lee. A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction. Chemom Intell Lab Syst. 2005;79:22–30.
31. Y Fu, U Kruger, Z Li, et al. Cross-validatory framework for optimal parameter estimation of KPCA and KPLS models. Chemom Intell Lab Syst. 2017;167:196–207.
32. TP Wright. Factors affecting the cost of airplanes. J Aeronaut Sci. 1936;3:122–128.
33. AB Badiru. Computational survey of univariate and multivariate learning curve models. IEEE Trans Eng Manag. 1992;39:176–188.
34. JA Hartigan, MA Wong. Algorithm AS 136: A K-Means Clustering Algorithm. J Royal Stat Soc Ser C (Applied Stat). 1979;28:100–108.
35. AK Jain, MN Murty, PJ Flynn. Data clustering: a review. ACM Comput Surv. 1999;31:264–323.
36. RJK Vegter, S de Groot, CJ Lamoth, DH Veeger, LH van der Woude. Initial skill acquisition of handrim wheelchair propulsion: A new perspective. IEEE Trans Neural Syst Rehabil Eng. 2014;22:104–113.
37. RJK Vegter, CJ Lamoth, S de Groot, DHEJ Veeger, LH van der Woude. Inter-individual differences in the initial 80 minutes of motor learning of handrim wheelchair propulsion. PLoS One. 2014;9:e89729.
38. P Jirapinyo, WM Abidi, H Aihara, et al. Preclinical endoscopic training using a part-task simulator: learning curve assessment and determination of threshold score for advancement to clinical endoscopy. Surg Endosc. 2017;31:4010–4015.
39. JA Dias, MF Dall'oglio, JR Colombo, RF Coelho, WC Nahas. The influence of previous robotic experience in the initial learning curve of laparoscopic radical prostatectomy. Int Braz J Urol. 2017;43:871–879.
40. CR Gallistel, S Fairhurst, P Balsam. The learning curve: Implications of a quantitative analysis. Proc Natl Acad Sci. 2004;101:13124–13131.
41. D Stefanidis, C Gardner, JT Paige, JR Korndorffer, D Nepomnayshy, D Chapman. Multicenter longitudinal assessment of resident technical skills. Am J Surg. 2015;209:120–125.
42. Y Hu, KD Brooks, H Kim, et al. Adaptive simulation training using cumulative sum: A randomized prospective trial. Am J Surg. 2016;211:377–383.
43. M Louridas, P Szasz, AB Fecso, et al. Practice does not always make perfect: need for selection curricula in modern surgical training. Surg Endosc. 2017;31:3718–3727.
44. TP Grantcharov, P Funch-Jensen. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. Am J Surg. 2009;197:447–449.
45. FHF Elneel, F Carter, B Tang, A Cuschieri. Extent of innate dexterity and ambidexterity across handedness and gender: Implications for training in laparoscopic surgery. Surg Endosc. 2008;22:31–37.
46. DT Hughes, SJ Forest, R Foitl, E Chao. Influence of medical students' past experiences and innate dexterity on suturing performance. Am J Surg. 2014;208:302–306.
47. AN Martin, Y Hu, IA Le, et al. Predicting surgical skill acquisition in preclinical medical students. Am J Surg. 2016;212:596–601.
48. AK Gardner, JM Marks, EM Pauli, A Majumder, BJ Dunkin. Changing attitudes and improving skills: demonstrating the value of the SAGES flexible endoscopy course for fellows. Surg Endosc. 2017;31:147–152.
49. T Nomura, T Matsutani, N Hagiwara, et al. Characteristics predicting laparoscopic skill in medical students: nine years' experience in a single center. Surg Endosc. 2018;32:96–104.
50. A Ali, Y Subhi, C Ringsted, L Konge. Gender differences in the acquisition of surgical skills: a systematic review. Surg Endosc. 2015;29:3065–3073.
51. H Mackenzie, AR Dixon. Proficiency gain curve and predictors of outcome for laparoscopic ventral mesh rectopexy. Surgery. 2014;156:158–167.
52. AP Berger, JC Giacalone, P Barlow, MR Kapadia, JN Keith. Choosing surgery as a career: Early results of a longitudinal study of medical students. Surgery. 2017;161:1683–1689.
53. PJ Roch, HM Rangnick, JA Brzoska, et al. Impact of visual–spatial ability on laparoscopic camera navigation training. Surg Endosc. 2018;32:1174–1183.
54. A Nemani, MA Yucel, U Kruger, et al. Assessing bimanual motor skills with optical neuroimaging. Sci Adv. 2018;4:eaat3807.
55. A Nemani, U Kruger, CA Cooper, SD Schwaitzberg, X Intes, S De. Objective assessment of surgical skill transfer using non-invasive brain imaging. Surg Endosc. 2019;33:2485–2494.
56. LP Sturm, JA Windsor, PH Cosman, P Cregan, PJ Hewett, GJ Maddern. A systematic review of skills transfer after surgical simulation training. Ann Surg. 2008;248:166–179.
57. Z Li, U Kruger, L Xie, A Almansoori, H Su. Adaptive KPCA modeling of nonlinear systems. IEEE Trans Signal Process. 2015;63:2364–2376.
58. C Hervas, PA Gutierrez, M Silva, JM Serrano. Combining classification and regression approaches for the quantification of highly overlapping capillary electrophoresis peaks by using evolutionary sigmoidal and product unit neural networks. J Chemom. 2007;21:567–577.
59. Y An, W Sherman, SL Dixon. Kernel-based partial least squares: Application to fingerprint-based QSAR with model visualization. J Chem Inf Model. 2013;53:2312–2321.