Peer Review File

Article information: https://dx.doi.org/10.21037/ls-22-29

Reviewer A:
Excellent job. Congratulations

<u>Responses to comments from Reviewer B:</u>

Comment 1: I would like to know how all the important events of laparoscopic learning curve can be explained accordingly to the 'Taxonomy of Learning Domains'? How the students should be validate in term of cognition, attitude and skill prospectives? Or, most of the teaching purpose we have came through in the article emphasized mostly in the willingness and innovation to teach but lack of realism method to construct scaling validation to what the students have learnt?

Reply 1: We agree that a discussion of the theoretical progression through the learning curve could add to our discussion. We have included an example using Bloom's taxonomy of learning. Unfortunately, this is not generally used for actual evaluation of a learner's progression on the learning curve for various reasons. We have edited the text to reflect this.

Changes in Text: As such, determining the factors by which to set the standard for a learning curve becomes important. From a theoretical standpoint, Bloom's taxonomy can be used to understand laparoscopic learning curves (48). This taxonomy contains six levels of learning, from the most basic (*remembering*) to the most complex (*creating*). In this framework, surgeons would first *remember* the steps of an operation and *understand* the purpose of each step. They would then *apply* those steps to perform an operation. After the operation, surgeons would *analyze* their results and *evaluate* their method to determine any improvements to their technique. Finally, they may *create* novel approaches to address challenges with the existing technique. While a useful model for understanding how a surgeon progresses along a learning curve cognitively, this is challenging to quantify rigorously and is not generally done. Resultantly, authors have evaluated measurable factors such as operative time, post-operative length of stay, estimated blood loss, conversion to an open procedure, post-operative complications, cancer recurrence, hospital readmission, and mortality when evaluating various procedures' learning curves (5,6,49–56).

Comment 2: Based on the 'establishing the learning curve' session, the statistical interpretation of the 'magic number' in term of different learning phases should be further explained.

Reply 2: We have updated the text to include further description of the "magic number." If it is felt that additional rigorous statistical discussion of standard or CUSUM or RA-CUSUM would be helpful to the readership of the journal, this could be included based on information from the Steiner paper from 2000 (see two relevant excerpts below).

2.1. Standard CUSUM

The CUSUM procedure is a well-established sequential monitoring scheme designed to detect changes in a process parameter of interest, denoted by, say, θ. The original formulation of the CUSUM is due to Page (1954). Two-sided implementations suggested by Barnard (1959) involved the use of a graphical device, called a V-mask. Unfortunately the V-mask is awkward to use in practice. An easier to use tabular form of the CUSUM can detect increases (or decreases) in θ. Using two tabular CUSUMs in conjunction accomplishes the goal of detecting any process changes. A standard tabular CUSUM involves monitoring

$$X_t = \max(0, X_{t-1} + W_t), \quad t = 1, 2, 3, \ldots, \tag{2.1}$$

where $X_0 = 0$, and W_t is the sample weight or score assigned to the tth subgroup. Subgroups are a collection of units taken from the production process at roughly the same time. Through a judicious choice of W_t the CUSUM can be designed to detect increases or decreases in θ. The CUSUM given by (2.1) sequentially tests the hypothesis $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_A$. The value of θ_0 is typically determined by the current process performance, while θ_A represents an alternate value of interest, corresponding typically to inferior performance. The process is assumed to be in state H_0 as long as $X_t < h$, and is deemed to have shifted to state H_A if $X_t \geq h$ at any time t. The constant h is called the control limit of the CUSUM. In quality-control terminology, a CUSUM that exceeds the control limit is said to have 'signalled'. A signal means that the chart has accumulated enough evidence to conclude that the process parameter has changed.

2.2. Risk-adjusted CUSUM

In most surgical contexts the risk of mortality estimated pre-operatively will vary considerably from patient to patient. An adjustment for prior risk is therefore appropriate to ensure that mortality rates that appear unusual and arise from differences in patient mix are not incorrectly attributed to the surgeon. We can adjust the CUSUM based on prior risk by adapting the magnitude of the scores using the patient's surgical risk, estimated pre-operatively. The surgical risk varies for each patient depending on risk factors present. We define $p_t(\theta) = g(\theta, \mathbf{x}_t)$, where $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})^T$ is a px1 vector reflecting the risk factors for patient t. The function g may be determined pre-operatively using a rating method such as Parsonnet risk factors (Parsonnet *et al.*, 1989), or may be based on a logistic regression model fitted to sample data. Since each patient has a different baseline risk level we define the hypotheses H_0 and H_A based on an odds ratio. Let R_0 and R_A represent the odds ratios under null and alternate hypotheses, respectively. To detect increases we set $R_A > R_0$. The choice of R_A is similar to defining the minimal clinically important effect in a clinical trial. If the estimated risk p_t is based on the current conditions we may set $R_0 = 1$. Given an estimated risk of failure equal to p_t, the odds of failure equals $p_t/(1 - p_t)$. Thus, for patient t under H_0 the odds of failure equals $R_0 p_t/(1 - p_t)$, whereas under H_A the odds of failure is $R_A p_t/(1 - p_t)$, which corresponds to a probability of failure equal to $R_A p_t/(1 - p_t + R_A p_t)$ under H_A. In this way the CUSUM repeatedly tests

$$\begin{aligned} H_0 &: \text{ odds ratio } = R_0 \text{ versus} \\ H_A &: \text{ odds ratio } = R_A. \end{aligned}$$

Then, the two possible log-likelihood ratio scores for patient t are:

$$W_t = \begin{cases} \log\left[\dfrac{(1 - p_t + R_0 p_t) R_A}{(1 - p_t + R_A p_t) R_0}\right] & \text{if } y_t = 1 \\ \log\left[\dfrac{1 - p_t + R_0 p_t}{1 - p_t + R_A p_t}\right] & \text{if } y_t = 0 \end{cases} \tag{2.3}$$

Changes in Text: To quantify the progress of a surgeon's training, many authors have sought to define laparoscopic procedures' learning curve. Prior authors have cited this as the "magic number" of cases, which signifies "the number of cases required to reach stability or technical competence" (5). This number will vary widely by surgeon based on the surgeon's prior experience in related open operations, unrelated MIS operations, and simulation (5).

Comment 3: As the training program requirements were outlined based on existing practical and theoretical knowledge, how can we upgrade the working capacity to the students? Can the learning curve assist establishing education program which fit the psychomotor domain layout (perception, set, guided response, mechanism, complex overt response, adaptation and origination)?
Reply 3: We have some discussion of how the learning curve may help with education, and have now included an additional discussion of this topic as noted below.
Changes in Text: Validation of existing and novel scoring systems, followed by expert consensus on the appropriate stepwise progression of training based on difficulty rating, can inform a standardized approach to laparoscopic skill progression that is replicable across institutions. This standardized approach can assist with learner feedback, case selection, and evaluation.

Comment 4: In the 'Training and Mentorship by Expert-Level Surgeons' session, the factor of how the rater assess reliability of a junior surgeons, as well as common errors and threats of validation should also be mentioned. For difficult procedure that only certain skillful experts handle, misinterpretation can easily be interfered by 'similar-to-me effect', for example. Moreover, how can we overcome the difficulty of 'how to train a trainer'? One who can operate beautifully doesn't mean he/she can also teach well, is there any training strategies for a mentor, as well as improving validation.
Reply 4: We have edited this section to include a brief discussion of the helpfulness of pioneer surgeons in pointing out errors in their successors. We have also included a mention of operative coaching, which is an extremely useful adjunct but a thorough discussion may be beyond our scope. We have included a reference to a recent systematic review of surgical coaching.
Changes in Text: This highlights the ability to shorten the phases of the surgical learning curve through training and mentoring via a master-apprentice model, e.g. early introduction to laparoscopic training in residency with structured mentorship from expert laparoscopic surgeons. Experts from the "pioneer" phase can provide tailored feedback to their successors based on prior experienced errors. Ideally, these experts would receive training in operative coaching to better refine learners' skills (64). For highly complex cases, further mastery can be achieved through completion of a dedicated fellowship.