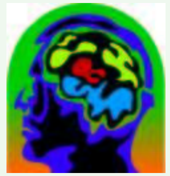
Writing posters efficiently: how to compare learning rates

Christine Lefebvre and Denis Cousineau

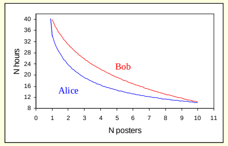
Université de Montréal

christine.lefebvre@umontreal.ca



Alice and Bob both learned to write posters. At first, preparing a poster took around 40 hours for both of them. At the end of training, both Alice and Bob were performing at an average preparation time of 12 hours. Does that mean both learned in the same way? If two different methods were used to train them, does that mean both were equally efficient? Not necessarily. In fact, if we look at their learning curves, we can see that Alice learned at a much faster rate than Bob. If the aim is to find the most efficient training method, then the learning rate is the critical value.

Figure 1 : Alice and Bob's learning curves



What is a learning curve?

The learning curve describes the change in performance throughout training. Data accumulated during training, for example response times, are plotted as a function of trials (or blocks of trials). The simplest form of learning curve would be a plot of the raw data as a function of trials (see left side of fig 2). However, this curve is usually very noisy. It can be better to plot summary values as a function of blocks of trials. For example, means and standard deviations are often used (see right side of figure 2). Data points can be fitted with an ideal curve, a learning curve. All learning curves have two scaling parameters in common. The first one is the asymptote (*a*) and the second one, amplitude (*b*). The identity of the curve is given by a « core function » often described by one parameter, the learning rate (*c*) (See eq. 1).

Eq 1 **: raw data functions**

Learning curve : $f(t) = a + b\, g(t)$

A → **Power curve** : $g_{PC}(t) = t^{c}$

B → **Exponential curve** : $g_{EX}(t) = e^{-ct}$

Eq 2 **: Averaged data functions**

Averaged leaning curve : $\bar{f}(n) = a + b\ \bar{g}(n)$

A → **Power curve** : $\bar{g}_{PC}(n) = \dfrac{n_F^{-(c-1)} - n_L^{-(c-1)}}{(c-1)N}$

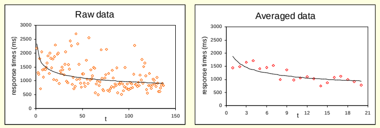
B → **Exponential curve** : $\bar{g}_{EX}(n) = \dfrac{e^{-c\,n_F} - e^{-c\,n_L}}{c\,N}$

Many models of automatization make predictions about the three parameters defining the learning curve. For example, Logan asserts that curvature (parameter c) should be equal for means and standard deviations. Cousineau and Larochelle's model makes the same prediction, while Rickart's extension of Logan's instance-based model predicts differing curvatures for means and standard deviations. One can also argue that strength theories and neural networks predict no change in curvature, but only in amplitude or asymptote throughout different tasks.

Fitting averages

Although learning is thought to occur on a trial by trial basis (Logan, 1988), learning curves are often plotted on data averaged by blocks of trials. Averaged data are used because they are less noisy than raw data. However, Heathcote and Mewhort (1995), and then Rickart (1997), pointed out that averaging raw data results in a different curve. For example, if raw data are best fitted with a power curve, the averaged data will no longer be. We show in figure 3 and equation 2 what is the resulting core function in case raw data core function is given. The example shown is for a power curve, but eq.2 shows how it can be done with an exponential curve. Fitting those averaged learning functions is easily achieved by a minimization algorithm. All methods proposed here allow the use of averaged or raw data (see Box 1). Therefore, it doesn't make a difference whether raw or averaged data are used when comparing curvatures.

Figure 2 **: Scatter plot of raw vs averaged data**



Box 1 **: fitting averages**

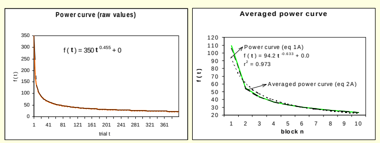
With **raw data** as a function of trial t

→ fit eq. 1

With **averaged data** as a function of block N

→ fit eq. 2 instead

Figure 3 **: Raw and averaged data fitted with a power function**



Advantages and limitations for each method

Method	Advantages	Limitations
Method A: **Estimation and comparison**	• Most obvious	•Weak because parameter estimation is slippery
Method B: **Assessment of proportionality**	• Powerful with restricted amount of data	•Limited power → Not recommended for less than 200 trials or 40 blocks of trials
Method C: **Normalizing**	• Simple to apply • No parameters estimation necessary → Not subject to biases	• Needs estimation of parameters a and b

Method A: Estimation and comparison

In order to test the predictions of the models cited above, one has to compare curvatures of learning curves. The most obvious method is to estimate the parameters from the data, and then compare the c parameters for each curve. The method is described in details in box 2.

Box 2 **: Estimation and comparison of c parameter**

•Minimization of Root mean square deviation or χ^2

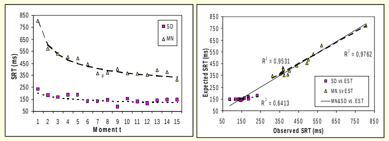
•Statistical tests on obtained estimated *c* parameters

→ example : coefficient of correlation of the observed data as a function of predicted data by estimated curve (Logan, 1988)

This method is flawed with two major biases. The first of these was identified by Heathcote and Mewhort (1995) who called it the « low asymptote bias ». They noticed, in Logan's (1988) data, that the estimated asymptotes seemed unrealistically low, even though the fit of the parameters with the data was very good. They found a non-linear relation between parameters a and c. The consequence of this relation is that a change in one of the two parameters will be compensated by a change in the other without any decrement of fit. The curvature can therefore be poorly estimated and invalidate the following comparison.

The second bias is called the « extended-range bias ». It occurs when Logan's method is applied to means and standard deviations (SDs), resulting in a linear regression in order to determine quality of fit. However, data of those two curves have by definition different ranges, SDs being smaller than means. As can be seen on the right side of figure 4, this extended range improves r^2 when both curves are plotted together rather than when plotted separately. So the correlation between the two curves is significant despite being different as can be seen on left side of figure 4.

Figure 4 **: Illustration of the extended-range bias**



Method B: Assessment of proportionality

These biases prompted the quest for alternative methods of comparing curvatures. The second method proposed here is easy to apply. As is demonstrated in box 3, if and only if two curves have the same curvature, they will be proportional. Plotting one against the other will result in a linear function. Box 4 illustrates how to apply the method.

Box 3 **: Assessment of proportionality**

If we create a scatter plot of f_1 as a function of f_2, we note that $f_2(t) = m\, f_1(t) + h$. Therefore,

$$m = \frac{\Delta f_2}{\Delta f_1} = \frac{f_2(t') - f_2(t'')}{f_1(t') - f_1(t'')} = \frac{b_2 g_2(t') + a_2 - b_2 g_2(t'') - a_2}{b_1 g_1(t') + a_1 - b_1 g_1(t'') - a_1} = \frac{b_2(g_2(t') - g_2(t''))}{b_1(g_1(t') - g_1(t''))}$$

The solution to this equation is a constant if :

1) there is no curvature in g_1 and g_2
→ Impossible if g_1 and g_2 are learning curves

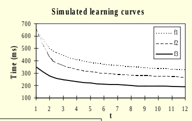
2) their difference is a constant
→ Impossible because *gs* do not have, by definition, an asymptote

3) g_1 and g_2 are in fact the same curve
→ The only possible solution

Box 4 **: Applying the proportionality method**

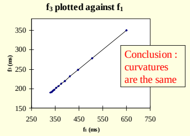
.Perform regression between the two data sets.
.Compute Durbin-Watson D statistics (available in most statistical packages) to test linearity of the function.

Parameters used to generate the curves	a	b	c
f_1	250	400	0,7
f_2	250	400	1,2
f_3	150	200	0,7

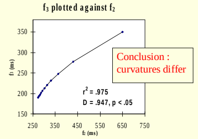


Note that parameters are not estimated (no low asymptote bias)

f_3 plotted against f_1



f_3 plotted against f_2



There is a limit to method B. Systematic deviation from linearity is very difficult to detect. Hence, the D test is not very powerful. Therefore, it is not recommended to use this method with less than 200 trials or 40 blocks of trials. Method C is not subject to that limit and can be used with small amount of data.

Method C : Normalizing curves

The normalization method is based on the assumption that if two curves have the same curvature, they should superimpose perfectly after normalization. Box 5 illustrates how to apply method C.

Box 5 **:Normalization of the curves**

• Estimate parameters a and b (visual inspection of the curves is recommended to detect inaccuracies in the normalizing process)

→ estimate a and b like shown for method A

or

→ assume the lowest data point to be the asymptote (vulnerable to outliers)

or

→ estimate the percent of fast errors, find the corresponding quantile and assume it to be the asymptote

• Proceed with the following transformation: $f_N(t) = \dfrac{f(t) - a}{b} = g(t)$, which isolates the core function

• **If summary values (mean or SDs)** are used, compute standard errors (SE)

→ If SE of both curves overlap for all points of the curve, then the two curves have the same curvature (but this is not a statistical test)

→ Perform an ANOVA of the normalized data with *type of curve* and *N* as variables.

Discussion

Two new methods for comparing learning curves have been described. They both offer interesting alternatives to the obvious but weak parameters estimation method. They can be used with both raw and averaged data. However, we recommend the use of averaged data; they are less noisy and more accurate tests can be performed on them.