Student Autonomy Improves Learning: A Theoretical Justification of the Empirical Results

Octavio Lerma and Vladik Kreinovich
Computational Science Program
University of Texas at El Paso
El Paso, Texas 79968, USA
lolerma@episd.org, vladik@utep.edu

Abstract

In many pedagogical situations, it is advantageous to give students some autonomy: for example, instead of assigning the same homework problem to all the students, to give students a choice between several similar problems, so that each student can choose a problem whose context best fits his or her experiences. A recent experimental study shows that there is a 45% correlation between degree of autonomy and student success. In this paper, we provide a theoretical explanation for this correlation value.

1 Formulation of the Problem

Traditional non-autonomous approach to learning. Traditionally, all the students in the class are assigned the same homework problems, the same practice problems, and are given the same problems on the tests.

Need for autonomy. The main advantage of the traditional approach seems to be that all the students are treated equally: they are given exactly the same problems, so they have the same chance to succeed.

In reality, however, the traditional approach has its limitations. For example, in a university setting, every engineering student has to take calculus. Engineering students usually form a majority in calculus classes. However, in addition to engineering students, also students from other disciplines, e.g., from bioinformatics, are required to take calculus. Since the majority of students in a calculus class are engineering students, most textbook application-related problems come from engineering or from related fields. This fact gives an unfair disadvantage to bioinformatics students many of whom are not very familiar with the main concepts from engineering.

As a result, it is much harder for a bioinformatics students to study in this class, and for those bioinformatics students who worked hard and mastered all the concepts, it is more difficult to show their knowledge on the tests – since the problems given on these tests are also skewed towards engineering-related topics.

This problem will not disappear if we simply switch to biology-related problems: then, it will be unfair to engineering students.

Similarly, in middle and high school, physics problems are often related to activities familiar to kids, such as soccer, basketball, skateboarding, etc. Such problems enliven the class but they give an unfair disadvantage, e.g., to female students who are usually less involved in soccer than boys.

In all these cases, to eliminate the unfairness of the traditional non-autonomous approach, it is desirable to give students some degree of *autonomy*. Specifically, on each assignment and on each test, instead of giving all the students exactly the same problem corresponding to a certain topic, it is desirable to give students *several* problems to choose from, so that each student will be able to decide which of the problem he or she wants to solve.

There is also an additional psychological advantage of providing students with autonomy: since the students themselves have to make choices, they feel more in control of the learning process, and it is well known that people usually perform better when they are (at least partly) in control than when they simply blindly follow others' instructions.

Autonomy indeed improves learning: an empirical fact. The need for student autonomy is well understood by many teachers, and many teachers have added elements of autonomy to their teaching. There is a large amount of anecdotal evidence showing that autonomy improves learning. Recently, this improvement was confirmed by a rigorous study [2] that showed that there is indeed a high correlation (45%) between the degree of autonomy and the students' success.

The empirical data needs a theoretical explanation: what we do in this paper. While the empirical data is very convincing, it is always desirable to come up with a theoretical explanation for this data. To be more precise, the paper [2] provides a deep *qualitative* theoretical explanation of its results. It is desirable to transform this qualitative explanation into a *quantitative* one, i.e., into an explanation that would not only explain the positiveness of the correlation, but that would also help us predict the numerical value of this correlation. Such an explanation is provided in this paper.

2 Formulation of the Corresponding Mathematical Model

Towards a model. To provide the desired quantitative explanation, let us formulate a simple mathematical model for autonomy.

As we have mentioned earlier, the main reason why we need autonomy is that for the same class of problems, at the same level of student understanding, the students will show different degree of success depending on how familiar they are with the overall context of this problem.

In other words, in different contexts, the students will exhibit different degrees of success x (e.g., grades). Let us fix a specific situation, i.e., a specific topic and a specific level of understanding of this topic. Let us denote, by $\underline{x}$, the smallest of these degrees corresponding to this level of understanding, and by $\overline{x}$, the largest of these degrees. Thus, depending on the context in which we present the corresponding problem, the grade of the same student may take any value from $\underline{x}$ to $\overline{x}$.

How to estimate probabilities of different values from $\underline{x}$ to $\overline{x}$? To properly gauge the effect of autonomy on student's learning, we should know the probability of different values x from the interval $[\underline{x}, \overline{x}]$.

In general, we have no reason to believe that some values from this interval are more frequent than others. So, it is reasonable to assume that all the values from this interval are equally probable, i.e., that we have a uniform distribution on this interval; see, e.g., [1].

How to describe case of autonomy. Let us now describe the probability distribution corresponding to the autonomy case.

Let us assume that, instead of single problem, we are given the student a choice between k different problems. For each of these problems, the success rate x_i $(i = 1, \ldots, k)$, the success rate is uniformly distributed in the interval $[\underline{x}, \overline{x}]$. Out of these k problems, the student will choose the one with which context he or she is most familiar, i.e., the one with the largest expected success rate $m_k \stackrel{\text{def}}{=} \max(x_1, \ldots, x_k)$. Thus, in the case of autonomy, the resulting success rate is distributed as the maximum of k independent random variables each of which is uniformly distributed on the given interval.

How many alternative problems should we design? On the one hand, the more choices, the better. On the other hand, good problems are not easy to design, and coming up with many additional problems would be very time-consuming. Let us therefore stop when the further increase in student success is not statistically significant.

Usually, in applications of statistics, a 5% threshold is used to describe statistical significance; see, e.g., [3]. So, we will stop when the difference between the expected grade $E[m_k]$ corresponding to k problems and the expected grade $E[m_{k+1}]$ (which will occurs if we add one more problem) does not exceed 5%.

For each k and for each x, the maximum m_k of k values x_i is smaller than or equal to x if and only if each of these values is $\le x$. Thus, due to independence assumption,

$$F_k(x) \stackrel{\text{def}}{=} \text{Prob}(m_k \le x) = \text{Prob}((x_1 \le x)\,\&\,\ldots\,\&\,(x_k \le x)) =$$

$$\text{Prob}(x_1 \le x) \cdot \ldots \cdot \text{Prob}(x_k \le x).$$

For the uniform distribution, $\text{Prob}(x_i \le x) = \dfrac{x - \underline{x}}{\overline{x} - \underline{x}}$, so $F_k(x) = \left(\dfrac{x - \underline{x}}{\overline{x} - \underline{x}}\right)^k$.

Thus, the corresponding probability density function $f_k(x)$ has the form

$$f_k(x) = \frac{dF_k(x)}{dx} = k \cdot \frac{(x - \underline{x})^{k-1}}{(\overline{x} - \underline{x})^k}.$$

Therefore, the mean grade $E[m_k]$ is equal to

$$E[m_k] = \int_{\underline{x}}^{\overline{x}} x \cdot f_k(x)\, dx = \int_{\underline{x}}^{\overline{x}} x \cdot k \cdot \frac{(x - \underline{x})^{k-1}}{(\overline{x} - \underline{x})^k}\, dx.$$

By introducing a new variable $x' \stackrel{\text{def}}{=} x - \underline{x}$, for which $x = \underline{x} + x'$, we can explicitly compute the corresponding integral, and get $E[m_k] = \underline{x} + \dfrac{k}{k+1} \cdot (\overline{x} - \underline{x})$. So, the $E[m_k]$ is at level $\dfrac{k}{k+1}$ in the interval $[\underline{x}, \overline{x}]$.

For $k = 1$, we get $\dfrac{1}{2} = 50\%$ of this interval. For $k = 2$, we get $\dfrac{2}{3} \approx 67\%$ of this interval – a statistically significantly larger value, since $67 - 50 > 5$. For $k = 3$, we get $\dfrac{3}{4} = 75\%$, which is also statistically significantly larger value ($75 - 67 > 5$). For $k = 4$, we get $\dfrac{4}{5} = 80\%$ which is *not* statistically significantly larger value, since $80 - 75 = 5$.

Thus, we select $k = 3$ alternatives for each problem. In this case, the probability distribution for the success rate m_3 can be described by the probability density $f_3(x) = k \cdot \dfrac{(x - \underline{x})^2}{(\overline{x} - \underline{x})^3}$.

Let us now compute the correlation. We have described the corresponding mathematical model. Let us now use this model to compute the correlation between the student's success and the autonomy level.

3 Analyzing the Mathematical Model

What we want to estimate. We want to find the correlation ρ between the success rate X and the autonomy level Y. In general, the correlation has the form

$$\rho = \frac{E[X \cdot Y] - E[X] \cdot E[Y]}{\sigma[X] \cdot \sigma[Y]}, \tag{1}$$

where $\sigma = \sqrt{V}$ is the standard deviation of the corresponding random variable.

How do we describe Y. Here, we only consider two levels of autonomy: no autonomy and giving a student the maximum choice (of 3 problems). Without losing generality, let us denote the autonomy case by $Y = 1$, and the non-autonomy case by $Y = -1$.

When researchers experimentally compare two techniques, they random assign each technique to some objects (in this case, to classes). To make comparison maximally fair, it is desirable to treat both techniques equally, in particular, to assign the same number of objects to each technique. In this case, we get $Y = 1$ and $Y = -1$ with the same probability 0.5. Thus, $E[Y] = 0$, so $V[Y] = E[(Y - E[Y])^2] = E[Y^2] = 1$ and $\sigma[Y] = \sqrt{V[Y]} = 1$.

Re-scaling X. It is known that the correlation does not change if we linearly re-scale each quantity x, i.e., replace it with $x' = a \cdot X + b$. For example, the correlation between height and weight should be the same whether we use inches and pounds or centimeters and kilograms.

We can use this fact to replace the original value $x \in [\underline{x}, \overline{x}]$ with an easier-to-analyze value $X = \dfrac{x - \underline{x}}{\overline{x} - \underline{x}} \in [0, 1]$. In this case, X is uniformly distributed on the interval $[0, 1]$ when $Y = 1$ and distributed as m_3, with probability density function $f_3(x) = 3x^2$, when $Y = -1$.

Estimating $E[X \cdot Y]$. Since both values $Y = \pm 1$ occur with probability $\dfrac{1}{2}$, we get:

$$E[X \cdot Y] = \frac{1}{2} \cdot E[X \cdot Y \mid Y = 1] + \frac{1}{2} \cdot E[X \cdot Y \mid Y = -1] =$$
$$\frac{1}{2} \cdot E[X \mid Y = 1] - \frac{1}{2} \cdot E[X \mid Y = -1].$$

For the uniform distribution, $E[X \mid Y = -1] = \dfrac{1}{2}$, and for the distribution m_3, as we have mentioned earlier, $E(X \mid Y = 1] = \dfrac{3}{4}$, so

$$E[X \cdot Y] = \frac{1}{2} \cdot \frac{3}{4} - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}. \tag{2}$$

Estimating $\sigma[X]$. Similarly, for the variance $V[X] = E[X^2] - (E[X])^2$, we have

$$E[X] = \frac{1}{2} \cdot E[X \mid Y = 1] + \frac{1}{2} \cdot E[X \mid Y = -1] = \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8},$$

and $E[X^2] = \dfrac{1}{2} \cdot E[X^2 \mid Y = 1] + \dfrac{1}{2} \cdot E[X^2 \mid Y = -1]$. For the uniform distribution, with uniform probability density $f_1(x) = 1$, we have

$$E[X^2 \mid Y = -1] = \int_0^1 x^2 \cdot f_1(x)\, dx = \frac{1}{3}.$$

Similarly,

$$E[X^2 \mid Y = 1] = \int_0^1 x^2 \cdot f_3(x)\, dx = \int_0^2 x^2 \cdot 3x^2\, dx = \frac{3}{5}.$$

Thus,

$$E[X^2] = \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{3}{5} = \frac{7}{15}.$$

Hence,

$$V[X] = E[X^2] - (E[X])^2 = \frac{7}{15} - \left(\frac{5}{8}\right)^2 = \frac{7}{15} - \frac{25}{64} = \frac{73}{15 \cdot 64}.$$

Hence,

$$\sigma[X] = \sqrt{V[X]} = \sqrt{\frac{73}{15}} \cdot \frac{1}{8}. \tag{3}$$

Resulting estimate. Substituting the expressions $E[Y] = 0$, $\sigma[Y] = 1$, (2), and (3) into the formula (1), we get

$$\rho = \frac{\frac{1}{8}}{\sqrt{\frac{73}{15}} \cdot \frac{1}{8}} = \sqrt{\frac{15}{73}} \approx 45\%.$$

Thus, we get a theoretical explanation for the empirical correlation.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. OCI-1135525 for the CI-Team Diffusion project: The Virtual Learning Commons and HRD-1242122 for the Cyber-ShARE Center of Excellence renewal.

References

[1] E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.

[2] J. Reeve and H. Jang, "What teachers say and do to support students' autonomy during a learning activity", *Journal of Education Psychology*, 2006, Vol. 98, No. 1, pp. 209–218.

[3] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC Press, Boca Raton, Florida, 2011.