

A simple mathematical model underlying Keller's individualized teaching method

Danilo T. Alves^{*1,2,3}, Nelson P.C. de Souza⁴, Silvio C.F. Pereira Filho⁵

¹Universidade Federal do Pará, Faculdade de Física, Belém, PA, Brasil.

²Universidade Federal do Pará, Instituto de Educação Matemática e Científica, Belém, PA, Brasil.

³Universidade do Minho, Centro de Física, P-4710-057, Braga, Portugal.

⁴Universidade Federal do Pará, Escola de Aplicação, Belém, PA, Brasil.

⁵Universidade Federal do Pará, Faculdade de Ciências Naturais, Campus Universitário do Marajó-Breves, Breves, PA, Brasil.

Received on September 23, 2021. Revised on June 20, 2022. Accepted on August 09, 2022.

Keller's article entitled "Good-bye teacher..." [J. Appl. Behav. Anal. 1, 79–89 (1968)] was fundamental for the development and dissemination of Keller's Personalized System of Instruction (PSI), which was one of the central issues in the discussions on psychology and education in the 1970s and 1980s, and nowadays has attracted attention in the context of the increasing use of online education. Belonging to a class of approaches usually named as mastery learning, PSI and modified PSI courses (Keller-type courses) present several interesting results, such as final grade distributions where the majority of students achieve the highest grades. Here, we present a simple mathematical model underlying Keller-type individualized teaching methods, describing, in terms of average characteristic parameters, the time evolution of the distribution of students per unit of content, and that most students achieve the highest grades at the end of the course. By applying this model to a real case of an introductory electromagnetism Keller-type course, we obtained its characteristic parameters with which we showed good agreement between the predictions and observations. The model presented here results in a simple formula, which is very accessible for use by a wide audience interested in planning or investigating Keller-type or other mastering learning methods.

Keywords: Keller method, personalized system of instruction, mastery learning, mathematical model.

1. Introduction

In his paper of 1968 [1], Keller describes his instructional plan (also named Personalized System of Instruction – PSI) as one through which the student can move at his own pace, showing mastery, not being forced to go ahead until he is ready. Like Bloom's learning for mastery [2–4], the PSI belongs to a class of approaches usually named as mastery learning. In Keller's scheme "...the student go ahead to new material only after demonstrating mastery of that which preceded, before to go to the next one" [1]. The mastery learning emerged as an alternative to traditional courses – those where students spend most of their time in classroom attending lectures, and assessments are only meant to score the success obtained at the course – and it has shown "positive effects on the examination performance of students in colleges, high schools, and the upper grades in elementary schools" [5]. In a Keller course, assessments are not only a measure of success, but are also formative. An assessment is considered to be formative if it "provides students opportunities to revise and hence improve the quality of their thinking and

learning" [6]. The formative aspect of the assessments in PSI is enhanced by the systematic feedback received by the students right after these assessments [1, 7]. Both formative assessments and systematic feedback are presently considered fundamental to learning [6].

The PSI originally started as a result of the application of the reinforcement theory, according to which the instruction process must be based on presentation, performance and consequences, maximizing the frequency of reinforcement and reducing the aversive consequences of errors [8, 9]. This led to the main PSI features, among them: mastery, self-pacing, linear small-step sequenced materials (each one related to a unit of content), repeated testing and immediate feedback [7]. In a Keller course, students spend most of their time in classroom doing assessments, receiving feedback or studying the instructional material: "instead of responding passively to a lecture, students must actively read, study, and respond in writing to questions over textual materials" [10]. In PSI, lectures are "...vehicles of motivation, rather than sources of critical information" [1] and are typically short [10]. An interesting result of the Keller method is the "production of a grade distribution that is upside down" [1], which means that, instead of a typical bell-shaped curve for the final grades found in

* Correspondence email address: danilo@ufpa.br

traditional courses [4], in PSI courses a higher percentage of students reaches the highest final grade.

Keller's method has been applied and discussed in the context of several disciplines [11], including Physics [12–16], and recent advances in the technology of information are being considered for implementing and improving the Keller course [17, 18, p. 201–221]. Although evidences derived from observations indicate success of the Keller scheme in many aspects, it began to decline in popularity during the 1980's. Among the causes, are criticisms on the behaviorist approach of the method, as well as the high workload associated with the implementation of a Keller course [17, 19].

In the present paper, we present a simple mathematical model to describe a PSI [1] or a modified PSI course (Keller-type course) [16]. With this model, we can estimate the evolution of the distribution of students per unit of content, the presence of the upside down effect in the grades and under which situations this effect occurs. We compare our theoretical predictions with results from a real application of a Keller-type method found in the literature [16].

The paper is organized as follows. In Sec. 2, we discuss a mathematical model to describe a very general Keller-type course. In Sec. 3, we propose approximations, obtaining a simple formula, which allows the visualization of several interesting aspects of Keller-type courses. In Sec. 4, we apply this model to a real case of an introductory level electromagnetism Keller-type course, described in Ref. [16]. In Sec. 5, we exemplify how our model can be used by a teacher interested in planning and improving a Keller-type course. In Sec. 6, we make our final considerations.

2. A General Model

In traditional courses, concepts and materials are divided into N_u units of content, which "... correspond, in many cases, to chapters in the textbook used in teaching" [4]. Usually, N_a (being $N_a \geq N_u$) tests are administered to the students. Usually each test covers the content of a given unit, so we can say that in traditional courses $N_a \approx N_u$. To the teacher, a test is "... an evaluation device that determines who learned those concepts well and who did not" [4]. To the students, each test means "... the end of instruction on the unit and the end of the time they need to spend working on those concepts" and most of the time "... the only chance to demonstrate what they learned" [4]. "After the test is administered and scored, marks are recorded in a grade book, and instruction begins on the next unit, where the process is repeated" [4]. The structure of a course according to the Keller plan [1] is very different and can be modeled as follows.

The course content is divided into N_u units, organized in a definite numerical order $(1, \dots, N_u)$, and the students have to show mastery in each unit by passing

assessments (tests, works, etc.). Let us consider that there are N_a total opportunities of such assessments related to these N_u units, so that, if the student fails to pass an assessment on the first opportunity, there can be other opportunities to be used. It can be noticed that in typical Keller courses $N_a \gg N_u$, whereas in traditional courses $N_a \approx N_u$. In addition, in a Keller course it can be applied N_e final examinations (but commonly $N_e = 1$), each one of them applied at the same time for all students. Usually, a certain percentage λ_u of the course grade is based on the number of units successfully completed during the term, a percentage λ_f is based on the final examinations, whereas the percentage $1 - (\lambda_u + \lambda_f)$ is associated with exercises (laboratory works, etc.). This means that there can have a minimal number N_c of units of content to be successfully completed by the students to get approval. Here, we call Keller-type to any course whose content is divided into N_u units, organized in a definite numerical order $(1, \dots, N_u)$, with the students having to show mastery in each unit before going to the next one, having a total of N_a opportunities of assessments, from which a number N_c of units of content must be successfully completed to get approval. The mathematical model presented here underlies Keller-type courses, which include the original Keller plan [1].

Let us consider a Keller-type course, with N_s students. In Fig. 1 we represent in a Cartesian plane the scheme of

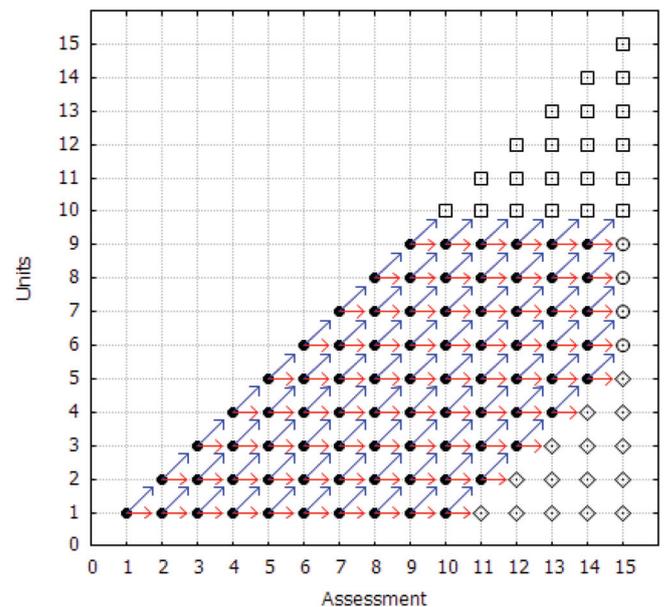


Figure 1: The scheme of assessments of a Keller course with $N_u = 9$, $N_c = 5$ and $N_a = 14$. In the horizontal axis we set a number for each opportunity of assessment [indicated by the first index, i , in $N_{(i,j)}$ along the text], whereas in the vertical axis we set a number for each unit of content [indicated by the second index, j , in $N_{(i,j)}$]. The points indicated by dark circles mean possible real situations of assessment. The points indicated by squares, diamonds and circles mean virtual situations.

assessments of a Keller course, with $N_u = 9$, $N_c = 5$ and $N_a = 14$. In the horizontal axis we set a number for each opportunity of assessment, whereas in the vertical axis we set a number for each unit of content. In this manner, the point (1,1) means that in the first opportunity of assessment the students can be trying to pass the assessment related to the first unit of content. The point (2,1) means that in the second opportunity of assessment, students are trying to pass the assessment related to the first unit of content, whereas the point (2,2) shows that in the same opportunity of assessment there are students trying to pass the assessment related to the second unit of content. The other points have analogous meaning. Considering $N_{(i,j)}$ as the number of students in the i th opportunity of assessment, trying to pass the assessment related to the j th unit of content, one can directly establish relations to describe the propagation of the number of students from a point to another, so that $N_{(i,j)}$ can be written as

$$N_{(i,j)} = \begin{cases} N_s & (i = j = 1), \\ \alpha_{(i-1,1)}N_{(i-1,1)} & (i \geq 2; j = 1), \\ \beta_{(i-1,i-1)}N_{(i-1,i-1)} & (i = j > 1), \\ \alpha_{(i-1,j)}N_{(i-1,j)} + \beta_{(i-1,j-1)}N_{(i-1,j-1)} & \\ (i \geq j + 1 > 2), \end{cases} \tag{1}$$

where i and $j = 1, 2, \dots$, the α coefficients are related to the failure in passing an assessment, whereas the β coefficients are related to the success (mastery in a given unit of content). In Fig. 1, the horizontal (red) arrows represent the propagation – mediated by the α coefficients – of the number of students that failed a given assessment, whereas the diagonal (blue) arrows represent the propagation – mediated by the β coefficients – of the number of students that obtain success. The points indicated by dark circles mean possible real situations of assessments, whereas the points indicated by squares, diamonds and circles are virtual points, which means that there are no real assessment related to these points. These virtual points are inserted because they will be useful (see Sec. 3) in the analysis of the time evolution of the number of students in each unit: those who have already concluded all real N_u units will “occupy” virtual units (squares); those students that have not concluded the minimal number N_c of units will occupy the diamond points; and those who concluded between N_c and N_u units occupy the circle points. This can be illustrated through the following examples: (i) students at the point (14,9) who pass this assessment complete all units, having no more assessments to pass, reach the “virtual” point (15,10); (ii) those at the point (14,9) and that do not pass this assessment go to the “virtual” point (15,9); (iii) and those at the point (14,5) and that pass this assessment reach the “virtual” point (15,6).

3. An Approximate Model

Equation (1) is very general, and a direct description underlying any Keller-type course. If all coefficients $\alpha_{(i,j)}$ and $\beta_{(i,j)}$ are known, equation (1) predicts the number of students $N_{(i,j)}$ in the i th opportunity of assessment, trying to pass the assessment related to the j th unit of content, in each moment of the course. The various elements that can affect such a course, for example, the number and quality of proctors and assistants, number of sessions, the quality of textbooks and other study materials, among others, can influence on a best (decreasing α and increasing β parameters) or worse (increasing α and decreasing β) student performance. Specifically, if students make more use of assessment opportunities, α decreases and β increases; on the contrary, taking less advantage of assessment opportunities, α increases and β decreases. In other words, all these characteristics can be encapsulated into the parameters $\alpha_{(i,j)}$ and $\beta_{(i,j)}$.

We point out that, in real situations, the use of equation (1) may involve a large number of parameters $\alpha_{(i,j)}$ and $\beta_{(i,j)}$, requiring some computational effort. Therefore, certain approximations with the intention of generating a more simplified model need to be considered. For example, a simplification is obtained by assuming that, for a given j th unit, the probability of success in passing does not depend on which is the i th opportunity of assessment used by the student [14]. In other words, $\beta_{(i,j)} = b_j$, where b_j is a value that only depends on j .

In order to obtain a simpler formula than those found in Ref. [14], but which already captures the essential features underlying Keller-type individualized teaching methods, and also can be used by a wide audience interested in planning or investigating Keller-type courses, here we introduce additional approximations. From now on we focus on describing our proposal for an approximate average behavior of $N_{(i,j)}$ given in equation (1). We consider that: (i) there is no dropping out of the course, so that the number N_s of students at any time will be always the same; (ii) in each opportunity, a given student try to pass in just one of the N_u units of content. Considering these approximations, we can write

$$\alpha_{(i,j)} + \beta_{(i,j)} = 1. \tag{2}$$

Hereafter, we will consider the approximate model where, for all i and j ,

$$\alpha_{(i,j)} = \alpha, \beta_{(i,j)} = \beta, \tag{3}$$

with α and β constants. This can be interpreted in two ways: the probability (or difficulty) of passing an assessment is exactly the same for all assessments (which can be considered unrealistic); α and β are average (taken over all units) characteristic parameters of a Keller-type course, and this is the interpretation we will adopt throughout the text.

Using equation (3) in equation (1), after some manipulations (see Appendix 6), we have

$$N_{(i,j)} = \frac{(i-1)! \alpha^{i-j} \beta^{j-1}}{(j-1)! (i-j)!} N_s, \tag{4}$$

where the factor multiplying N_s is the Bernoulli probability function, and, as already mentioned, $N_{(i,j)}$ is the number of students at the position (i, j) of the diagram in Fig. 1.

To illustrate an use of equation 4, let us consider, for instance, the case $(\alpha, \beta) = (1/2, 1/2)$. For this situation, we get the time (given in terms of the number of assessments already done) evolution of the class as shown in Fig. 2, which exhibits the situations of the class after the 4th, 9th, 14th and 19th assessments. For instance, after 4 assessments (blue line with points indicated by diamonds), the situation is described by $N_{(5,j)}/N_s$, and the figure shows that $N_{(5,1)}/N_s \approx 6\%$ of the students need to redo the assessment of the 1st unit, whereas $N_{(5,2)}/N_s \approx 25\%$ are able to do the assessment of the 2nd unit, $N_{(5,3)}/N_s \approx 38\%$ are able to do the assessment of the 3rd unit, $N_{(5,4)}/N_s \approx 25\%$ of the 4th unit, and $N_{(5,5)}/N_s \approx 6\%$ are able to do the assessment of the 5th unit. The red line (with points indicated by squares) shows that after the 9th assessment and for $N_u = 9$, among other results, $N_{(10,9)}/N_s \approx 1.8\%$ of the students are able to do the assessment of the last 9th unit and $N_{(10,10)}/N_s \approx 0.2\%$ are at the “virtual” unit 10, which means that these students have already shown mastery in all 9 units. The purple line (with points indicated by crosses) shows, after the 19th assessment and for $N_u = 9$, among other results, that the sum over the virtual units $\sum_{j=10}^{20} N_{(20,j)}/N_s \approx 67.6\%$ gives the part of the students that have already shown mastery in all the

$N_u = 9$ units of content. The lines blue (diamonds), red (squares), green (triangles) and purple (crosses) exhibit the behavior of a “wave packet” propagating from the left to right in Fig. 2. For other values of α and β , we get the same behavior, but as β increases (it becomes easier to pass assessments) the peak of the wave packet propagates faster, showing that students obtain a faster progress (mastery) in the units of content.

To investigate the production in Keller courses that is upside down [1], let us consider the hypothesis that as the number of units of content in which a student shows mastery increases, the probability of achieving the best final grades also increases. Then, we assume that the behavior of final grades is intrinsically related to the mastery level reached by the students. Let us consider, for example, the model described by equation (4) with $(\alpha, \beta) = (1/2, 1/2)$, and also $N_u = 9$ for the following situations: when $N_a = 9, N_a = 14$ and $N_a = 19$. In Fig. 3 we observe the distribution of students versus the maximum number of the units of content for which they obtained mastery at the end of the course. For instance, the squares on the red line show that for $N_a = 9$ we have a distribution of mastery which resembles a bell-shaped curve, which is typical of traditional courses [4]. We interpret this proximity between the results for mastery in a Keller-type course and traditional courses as a consequence of our choice of a theoretical model of a Keller scheme with $N_a = N_u$, being the relation $N_a \approx N_u$ typical in traditional courses [4]. However, in real Keller-type courses $N_a > N_u$ and one can observe in Fig. 3 that the distribution of mastery suffers an inversion as N_a becomes greater than N_u . For $N_a = 14$ (triangles on the green line), we already observe a deformation of the “bell-shaped” curve. For $N_a = 19$ (crosses on the purple line), we finally observe an inversion: a transition from a bell-shaped to an “exponential-shaped” curve, indicating

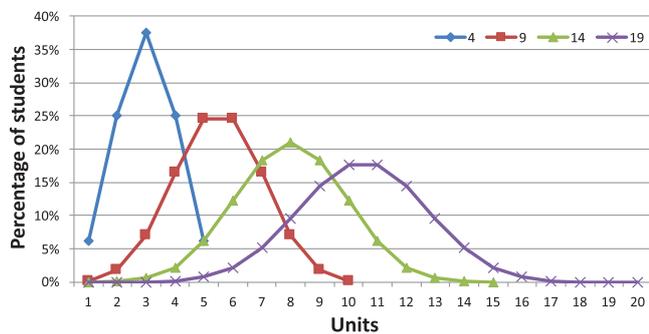


Figure 2: For the case $(\alpha, \beta) = (1/2, 1/2)$, the horizontal axis gives the number of each unit of content, whereas the vertical axis exhibits the percentage of students ($N_{(i,j)}/N_s$) who are able to do the assessment correspondent to each unit. After 4 assessments, the situation is described by the blue line (diamonds), which shows $N_{(5,j)}/N_s$ for $j = 1..5$; after the 9th assessment, the red line (squares) shows $N_{(10,j)}/N_s$ for $j = 1..10$; after the 14th assessment, $N_{(15,j)}/N_s$ for $j = 1..15$ is described by the green line (triangles); and after the 19th assessment, the situation is described by $N_{(20,j)}/N_s$ for $j = 1..20$, which is indicated by the purple line (crosses).

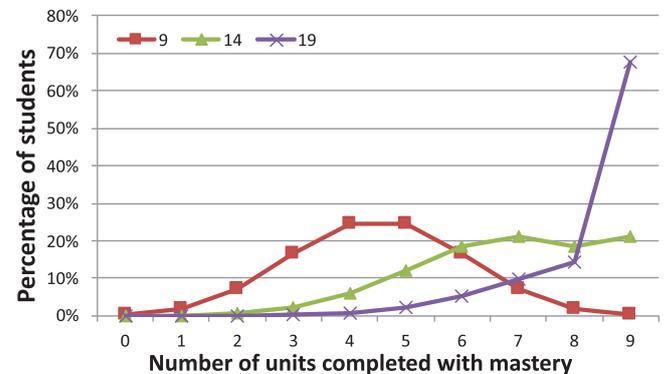


Figure 3: The horizontal axis shows the number of units completed with mastery. The vertical axis exhibits the percentage of students. The cases considered are defined by $(\alpha, \beta) = (1/2, 1/2)$, $N_u = 9$ and for the following values of N_a : for $N_a = 9$ (squares on the red line); for $N_a = 14$ (triangles on the green line); for $N_a = 19$ (crosses on the purple line).

that the majority of students obtained mastery in all $N_u = 9$ units. Then, assuming that the behavior of final grades is related to the mastery level reached by the students, an inversion of the mastery curve can be directly mapped into an inversion of final grades, which corresponds to the upside down effect mentioned by Keller [1]. But, we remark that the present model reveals that this inversion does not occur for an arbitrary value of N_a , but, as the number N_a increases in comparison to N_u , the possibility of appearance of the mentioned inversion effect enhances.

4. Application to a Real Keller-type Course

Now, we apply equation 4 to a real case of an introductory electromagnetism Keller-type course, described in Ref. [16]. In Fig. 4, we exhibit the wave packet behavior (diamonds on the blue line) found in the real situation (discarding waivers) described in Ref. [16]. We also exhibit the prediction from our model (triangles on the green line), considering the average characteristic parameters $(\alpha, \beta) = (2/5, 3/5)$. One can see that the real behavior is in good agreement with the model, for these characteristic parameters. This indicates that $(\alpha, \beta) = (2/5, 3/5)$ are the average characteristic parameters for the Keller-type course described in Ref. [16].

Let us use the average characteristic parameters $(\alpha, \beta) = (2/5, 3/5)$ to investigate the mastery. In Fig. 5, we exhibit the final mastery situation, considering $(\alpha, \beta) = (2/5, 3/5)$, and compare the theoretical prediction using these parameters (triangles on the green line) with the real situation (diamonds on the blue line) described in Ref. [16], where it was considered, effectively, $N_u = 11$ and $N_a = 16$. We highlight that the inversion of grades is in good agreement with the theoretical prediction. This reinforces that

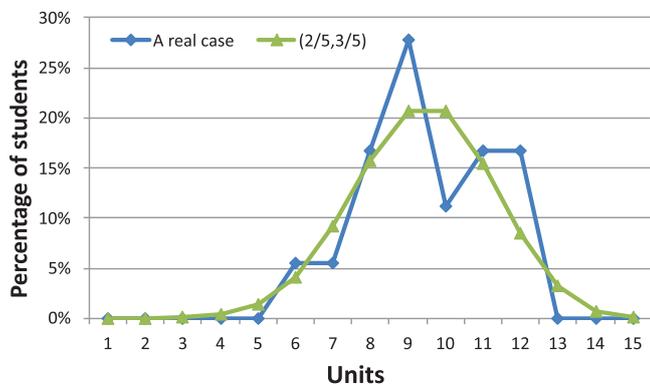


Figure 4: The horizontal axis shows the number of each unit of content, whereas the vertical axis exhibits the percentage of students ($N_{(i,j)}/N_s$) who are able to do the assessment correspondent to each unit, after the 14th assessment, for the case $(\alpha, \beta) = (2/5, 3/5)$ (triangles on the green line), and a real case described in Ref. [16] (diamonds on the blue line).

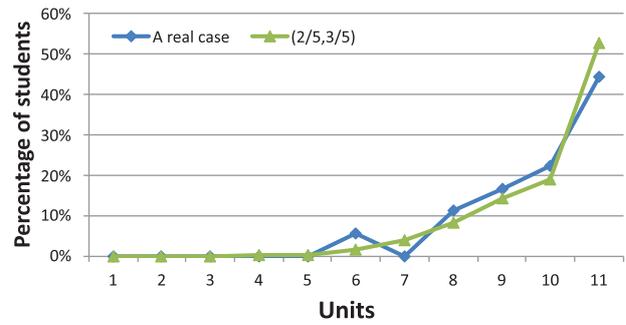


Figure 5: The horizontal axis shows the number of units completed with mastery. The vertical axis exhibits the percentage of students. The cases considered are $(\alpha, \beta) = (2/5, 3/5)$ (triangles on the green line), and a real case described in Ref. [16] (diamonds on the blue line). We considered $N_u = 11$ and $N_a = 16$.

$(\alpha, \beta) = (2/5, 3/5)$ are the average characteristic parameters of the Keller-type course considered in Ref. [16].

5. Use in Planning and Improving a Keller-type Course

Let us exemplify how equation (4) can be used, for example, by a teacher interested in planning and improving a Keller-type course. Among several characteristics that may differ from one Keller-type course to another, in common is the need to divide the content into N_u units, and define the number N_a of total assessment opportunities. If, for example, the teacher divides the content into $N_u = 9$ units, and, from previous experience, estimates that an average (over the units) of 50% of students will have success in passing each assessment, the parameters characterizing such a course are $(\alpha, \beta) = (1/2, 1/2)$. A question for the teacher is: how many assessment opportunities are needed for most students to reach mastery at the end of the course?

Using equation (4) (as explained in Section 3), for $N_a = N_u = 9$, the mastery curve will be a bell-shaped curve found in traditional courses (see the squares on the red line in Fig. 3) [4]. On the other hand, if the choice is $N_a = N_u = 19$, the teacher can expect a transition from a bell-shaped to an “exponential-shaped” curve, indicating that the majority of students obtain mastery in all $N_u = 9$ units (see the crosses on the purple line in Fig. 3), which is a typical result of a Keller-type course [1, 16].

Let us now consider that, after the application of a Keller-type course, the data collected by the teacher show that the class did not exhibit the evolution predicted in Fig. 2, or the mastery curve predicted in Fig. 3, both considering $(\alpha, \beta) = (1/2, 1/2)$. Then, from this, the teacher could find the parameters (α, β) that are in agreement with the results obtained in the applied Keller-type course, which allows an improvement in the choice of the number N_a of total assessment opportunities, which would be useful in other applications. Thus,

equation (4) can be used for planning and improving a Keller-type course

6. Final Remarks

Our main result is the simple formula in equation (4), which predicts the average behavior of the number of students $N_{(i,j)}$ in the i th opportunity of assessment, trying to pass the assessment related to the j th unit of content, in each moment of a Keller-type course. $N_{(i,j)}$ is described in terms of only two parameters, the coefficient α , which is the average probability of failing in passing an assessment, and β , which is the average probability of success (or to get mastery in a given unit of content). The various characteristics of a Keller-type course, for example, the number and quality of assistants, number of sessions, the quality of textbooks and other study materials, among others, can influence on a best (decreasing α and increasing β parameters) or worse (increasing α and decreasing β) student performance, with all these characteristics encapsulated in the α and β parameters. As discussed throughout the text, this makes equation (4) very accessible for use by a wide audience interested in planning or investigating Keller-type courses or other mastering learning courses.

An interesting result coming from our analysis based on equation (4) is that when the number N_a of assessment opportunities is approximately the same as the number N_u of units of content, the mastery curve will be a bell-shaped curve found in traditional courses (red line in Fig. 3), whereas as N_a becomes greater than N_u , there is a change from a bell-shaped to an exponential-shaped curve (purple line in Fig. 3), indicating that the majority of students obtain mastery, which is a typical result of a Keller-type course.

Acknowledgments

We thank Liane R. Tarouco, Adilson O. do Espírito Santo, Olavo F. Galvão, Benedito de Jesus P. Ferreira, César A. Galera, Tadeu O. Gonçalves, Ana Leda de F. Brino, Ícaro Gonzaga, Romariz S. Barros and Thiago D. da Costa, for valuable discussions. This work was partially supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brazil).

Supplementary material

The following online material is available for this article:

Appendix A – Obtaining equation (4)

This section builds on using equation 1 in equation 2. For $i = 2$ and $j = 1, 2$, we get

$$N_{(2,1)} = \alpha N_s \quad (5)$$

$$N_{(2,2)} = \beta N_s. \quad (6)$$

In parallel, note that we can write

$$(\alpha + \beta)N_s = \alpha N_s + \beta N_s, \quad (7)$$

which, for convenience, can be written as

$$(\alpha + \beta)^{2-1}N_s = \underbrace{\alpha^{2-1}\beta^{1-1}N_s}_{N_{(2,1)}} + \underbrace{\alpha^{2-2}\beta^{2-1}N_s}_{N_{(2,2)}}. \quad (8)$$

From equation (8), we identify $N_{(2,1)}$ and $N_{(2,2)}$.

For $i = 3$ and $j = 1, 2, 3$, we have

$$N_{(3,1)} = \alpha^2 N_s, \quad (9)$$

$$N_{(3,2)} = 2\alpha\beta N_s, \quad (10)$$

$$N_{(3,3)} = \beta^2 N_s. \quad (11)$$

In parallel, we can also write

$$(\alpha + \beta)^2 N_s = \alpha^2 N_s + 2\alpha\beta N_s + \beta^2 N_s, \quad (12)$$

which can be conveniently written as

$$\begin{aligned} (\alpha + \beta)^{3-1}N_s &= \underbrace{\alpha^{3-1}\beta^{1-1}N_s}_{N_{(3,1)}} + \underbrace{2\alpha^{3-2}\beta^{2-1}N_s}_{N_{(3,2)}} \\ &+ \underbrace{\alpha^{3-3}\beta^{3-1}N_s}_{N_{(3,3)}}. \end{aligned} \quad (13)$$

From equation (13), we identify $N_{(3,1)}$, $N_{(3,2)}$, and $N_{(3,3)}$.

For $i = 4$ and $j = 1, 2, 3, 4$, we have

$$N_{(4,1)} = \alpha^3 N_s, \quad (14)$$

$$N_{(4,2)} = 3\alpha^2\beta N_s, \quad (15)$$

$$N_{(4,3)} = 3\alpha\beta^2 N_s, \quad (16)$$

$$N_{(4,4)} = \beta^3 N_s. \quad (17)$$

We can write

$$(\alpha + \beta)^3 N_s = \alpha^3 N_s + 3\alpha^2\beta N_s + 3\alpha\beta^2 N_s + \beta^3 N_s, \quad (18)$$

which, for convenience, can be written as

$$\begin{aligned} (\alpha + \beta)^{4-1}N_s &= \underbrace{\alpha^{4-1}\beta^{1-1}N_s}_{N_{(4,1)}} + \underbrace{3\alpha^{4-2}\beta^{2-1}N_s}_{N_{(4,2)}} \\ &+ \underbrace{3\alpha^{4-3}\beta^{3-1}N_s}_{N_{(4,3)}} + \underbrace{\alpha^{4-4}\beta^{4-1}N_s}_{N_{(4,4)}}. \end{aligned} \quad (19)$$

From equation (13), we identify $N_{(4,1)}$, $N_{(4,2)}$, $N_{(4,3)}$, and $N_{(4,4)}$.

The same pattern found in equations (8), (13), and (19) can be obtained for $(\alpha + \beta)^s$, with $s = 4, 5, 6, \dots$. Thus, from these equations, one can see that $N_{(i,j)}$ can be written in terms of binomial coefficients. It is well known that

$$(\alpha + \beta)^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} \alpha^k \beta^{n-k}. \quad (20)$$

Taking as basis equations (8), (13), and (19), and making

$$n = i - 1, \quad (21)$$

$$k = i - j, \quad (22)$$

equation (20) can be written as

$$(\alpha + \beta)^{i-1} = \sum_{j=1}^i \frac{(i-1)!}{(i-j)!(j-1)!} \alpha^{i-j} \beta^{j-1}. \quad (23)$$

Multiplying both sides of equation (23) by N_s , and comparing with equations (8), (13), and (19), we can write

$$(\alpha + \beta)^{i-1} N_s = \sum_{j=1}^i \underbrace{\frac{(i-1)!}{(i-j)!(j-1)!} \alpha^{i-j} \beta^{j-1}}_{N_{(i,j)}} N_s, \quad (24)$$

from which we identify $N_{(i,j)}$, and thus obtaining equation (4).

References

- [1] F. Keller, *Journal of Applied Behavior Analysis* **1**, 79 (1968).
- [2] S. Benjamin, E. Dhew and B. Bloom, *Eval. Comment* **1**, 1 (1968).
- [3] B.S. Bloom, *Educational Researcher* **13**, 4 (1984).
- [4] T.R. Guskey, *Implementing Mastery Learning* (Wadsworth Publishing Company, Belmont, 1997), 2nd.
- [5] C.C. Kulik, J.A. Kulik, R.L. Bangert-Drowns and R.E. Slavin, *Review of Educational Research* **60**, 265 (1990).
- [6] J.D. Bransford, A.L. Brown and R.R. Cocking, *How People Learn: Brain, Mind, Experience and School: Expanded Edition* (The National Academies Press, Washington, 2000).
- [7] J.G. Sherman, *Journal of Applied Behavior Analysis* **25**, 59 (1992).
- [8] B.F. Skinner, *Science and Human Behavior* (MacMillan, New York, 1965).
- [9] B.F. Skinner, *Harvard Educational Review* **24**, 86 (1954).
- [10] W. Buskist, D. Cush and R.J. DeGrandpre, *Journal of Behavioral Education* **1**, 215 (1991).
- [11] J.A. Kulik, C.L. Kulik and K. Carmichael, *Science* **183**, 379 (1974).
- [12] M.A. Moreira, *Revista Brasileira de Física* **3**, 157 (1973).
- [13] P.C. Bezerra and L.C. Gomes, *Revista Brasileira de Física* **3**, 139 (1973).
- [14] P.C. Bezerra, L.C. Gomes and J. Mendes Filho, *Revista Brasileira de Física* **4**, 175 (1974).
- [15] P.H. Dionisio and M.A. Moreira, *Revista Brasileira de Física* **5**, 131 (1975).
- [16] D.T. Alves, S.A. Souza, S.C. Filho and W. Elias, *Revista Brasileira de Ensino de Física* **33**, 1403 (2011).
- [17] H.L. Eyre, *Behavior Analyst Today* **8**, 317 (2007).
- [18] E.J. Fox, in: *Evidence-based Educational Methods*, edited by D.J. Moran and R.W. Mallot (Elsevier Academic Press, San Diego, 2004).
- [19] R. Silberman, *Journal of Chemical Education* **55**, 97 (1978).