A SYSTEMS APPROACH TO EXAMINATION

R. F. VAN NAERSEN

SUMMARY

For a systems theoretical approach to examination in higher education, a simple mathematical model is proposed in which 'knowledge' relates linearly to the proportion of items answered correctly. The increase of knowledge, if forgetting is ignored, is proportional to 'capacity' and 'engagement' and inversely proportional to the 'extent' of the subject matter. Continual forgetting is proportional to the already obtained knowledge and to the 'isolatedness' of the subject matter, and inversely proportional to 'memory'. When a student fails an examination he loses time. It is assumed that the student follows a strategy in which he minimizes the expectation of his total effort. In this case it is possible in principle to organize the educational system so that, given the requirements of minimum sufficient scores and given the subject matter, the velocity of the stream of students is at a maximum. The use of the model is illustrated with two examples. First, calculation of the necessary increase of the cutoff score and the lengthening of the test if the number of examinations per year has to be enlarged and the optimal knowledge level is not allowed to become lower nor the expected study time to increase. Second, comparison of the expected study time in case of (1) undivided subject matter, (2) conjunctive combination of two parts of the subject matter, and (3) compensatory combination of the same two parts. The calculation of the compensatory combination had to be done with a Monte-Carlo method.

1. MINIMIZING EXPECTED STUDY TIME

In the Netherlands, as in many other continental countries, the university student has to pass a large number of examinations. When he has passed them all, he receives his diploma. In the arrangement of his study, for example in the time he spends on every examination, and even the sequence of the examinations, he has much freedom. The consequence, of course, is a large dispersion of study time. The efficiency of this free system, which has certain advantages over a method with a rigid rate of study, may be enhanced by systems theoretical principles.

It is possible to maximize the total knowledge of the students with a constant study time. We will, however, adhere to the existing system of fixed requirements, which has the advantage of a recognizable diploma of constant value. With these requirements the total utility
of the university, the society, and the student can be maximized, but one of the most important parts of this total utility may be the value of the student's time. We will restrict ourselves to this utility assuming that the educational effort of society is constant. Thus, given the requirements of the examinations and the educational effort in the shape of books, laboratories, lectures, and the like, the student's total time to pass all the examinations will be minimized.

One has to take account of the phenomenon of failing examinations. In the Netherlands this does not have the fatal consequence it may have in other countries. The student may try again some months later, and if he fails again he may continue trying; he only loses time. By preparing himself longer before the examination the student can decrease the chance of failing, but this probability is never zero, even if the student knows all the answers, because he can make mistakes; every score has a measurement error. This means that there must be an optimal effort before an examination, in such a way that the expectation of the total study time, only meaning the time the student spends studying for the examination, is minimized. The level to which the student prepares himself may be called his strategy, the variable which he uses to minimize the expectation of his total effort. Parts of the subject matter, the time between examinations, the method of combining scores on different examinations, and the length (number of items) of examinations may be varied. In the system all these conditions and values can be varied until an optimal solution is found in which the total time that a group of students needs to pass a sequence of examinations is minimized.

A solution is possible only if it is assumed that the student adopts a strategy. We will assume that the student follows the optimal strategy in choosing a level of knowledge. He is able to do so because he knows or estimates his true score \( t \).

2. CURVES OF LEARNING AND FORGETTING

When drafting a model simplifications of reality are inevitable. Later empirical research must demonstrate the 'reality' of the model. The first simplification is that the subject matter is seen as a set of questions of which the student must know the answers, or as a set of behaviours he has to learn. The requirement of the examinations is the proportion of the set that must be known by the student. The examination itself is a random selection of the domain of questions. The length, the number of items, of the examination is \( k \), of which \( y \) must be answered correctly.

The most important variable of the model is the ability \( t \) of the student, defined as the proportion of the items in the domain which the student can answer correctly; \( t \) is the true score in the domain sampling model of test theory. When a multiple choice item is concerned, \( t \) will be corrected for guessing with the formula

\[
t' = t_0 (1 + \frac{I - t_0}{t_0})
\]

in which \( t_0 \) is the proportion of items in the domain which can be answered correctly by guessing. In two-choice items \( t_0 \) is 0.5 or a little more, in four-choice items 0.25 or somewhat larger, and so on; \( t \) will be called knowledge.

Reflection suggests that during study knowledge \( t \) increases with time, and per time unit with learning capacity for the subject matter concerned, and with the intensity of learning, which we will call engagement. Knowledge \( q \), increases less when the extent of the subject matter is larger. Length and difficulty can be distinguished when considering extent of subject matter, but this distinction will not be needed in the following discussion. The student continually forgets part of the subject matter, and this forgetting, the decrease of \( t \), will increase with the already learned quantity. It also depends on the memory of the student and on the coherence of the subject matter. We will speak about the memory \( m \) and the isolatesness \( i \) of the subject matter. We will express the time in weeks, \( w \), and call the velocity of learning or forgetting \( v \). The velocity is the increase of the corrected true score (the knowledge \( t \)) with time.

The model assumes a simple relation between \( v \) and the parameters of the student \( g \), \( c \) and \( m \), and of the subject matter \( x \) and \( i \). It is postulated that (1) if there were no forgetting the learning velocity would be inversely proportional to the extent of the subject matter \( x \), and proportional to the engagement \( g \) and the capacity \( c \) of the student; and (2) forgetting is proportional to the knowledge \( t \), and the isolatesness \( i \), and inversely proportional to the memory \( m \); the forgetting velocity can be added to the learning velocity and a person also forgets during learning.

\[
v = c \frac{d}{dw} \frac{d}{dw} = g \frac{c}{x} \frac{i}{m} \frac{d}{dw}
\]

(2)

Solution of this differential equation with the boundary condition that the knowledge at some initial point be \( t_0 \), gives

\[
t = g \frac{c}{x} (1 - \frac{g}{x} \frac{c}{x} \frac{m}{t_0} \frac{e}{w})
\]

(3)

The initial learning curve arises from this by putting \( t_0 = 0 \):

\[
t = g \frac{c}{x} (1 - \frac{e}{w})
\]

(4)

The first derivative of this learning function to the time \( w \), is always positive and the second derivative always negative, as is necessary for a learning curve. Moreover it can be seen that exact mastery can be reached only if \( g/m/xi \) is greater than 1, and passing the examination is improbable if \( g/m/xi \) is smaller than \( y \), corrected for guessing and divided by \( k \). This means that in the model, just as in reality, not every student can pass every examination, even if he tries very hard, because \( g \) has an upper limit. The more difficult, longer, and incoherent the subject matter, the greater the product \( xi \) and the more the examination becomes selective. The parameter \( i \), however, can be decreased with better instruction, for instance if more relations between parts of the subject matter are shown.

By putting \( g \) in (3) equal to zero one gets the 'forgetting curve':
\[
t_0 = t_b e^{-w t/m}
\]

in which \( t_b \) may be the knowledge on the day of examination and \( w \) the number of weeks thereafter. The knowledge decreases exponentially, as in many learning models. From (5), \( m \) can be solved:

\[
m = \frac{w}{\ln(t_0/t_b)}
\]

An arbitrary subject can be called "standard", that means the \( x \) and the \( i \) of the subject matter are put equal to one. With (6), \( m \) can be found empirically by measuring the knowledge at the examination and \( w \) weeks later, when knowledge is decreased from \( t_b \) to \( t_0 \). When the \( m \) of the person is known, the capacity can be found empirically with (3) or (4). By administering another examination to the same person(s), the \( x \) and \( i \) of the new subject matter can be calculated.

In every formula \( c/x \) and \( m/i \) can be treated as two single parameters; \( x \) and \( i \), however, are maintained to remind the reader that not only personal parameters are relevant but also subject matter parameters. One can understand \( c/x \) to be the proportion of the subject matter that can be learned in a week if \( g = 1 \) (usually maximum learning) and if forgetting can be ignored; \( m/i \) can be understood as \( 1.44 = 1/\ln 2 \) times the number of weeks in which knowledge halves if \( g \) is zero (no learning).

3. PROBABILITY OF SUCCESS AND TOTAL EFFORT

With the assumption that the score of the person at the examination (of \( k \) items of which \( y \) must be answered correctly) given the true score \( t \), is distributed binomially, the probability of success is

\[
p = \sum_{j=y}^{k} \binom{k}{j} t^j (1-t)^{k-j}
\]

The concept effort can be defined as the product of engagement and time. In the model the student minimizes total effort \( f \):

\[
f = \sum_{j} e_j w_j
\]

The effort \( f \) can be expressed in weeks of, say, forty hours. In this case normally has values between 0 and 1. For a student who studies sixty hours a week intensively for an examination, \( g \) would be 1.5, but he cannot maintain that rate. For a student who works thirty hours a week and who divides his attention equally over three examinations \( g \) would be 0.25.

As already noted, the choice of \( t_0 \) is the 'strategy' of the student. If he chooses \( t_0 \) too low, the probability of failing, \( q = 1 - p \), becomes too large. On the other hand, a (too) high \( t_0 \) means an unnecessary amount of preparation time. There may be an optimal knowledge, \( t_0 \), with which the expectation of the total effort, \( E(f) \), is minimal.

Let \( f_0 \) be the effort necessary to increase the knowledge from zero to \( t_{\text{max}} \), the knowledge on the day of examination, and \( f_2 \) be the effort necessary after each failure and forgetting period to reach \( t_{\text{max}} \) again; let the time interval between successive examinations be constant \( w_2 \), then

\[
E(f) = f_0 + q f_1 + q^2 f_2 + \cdots = f_0 + \frac{1 - \frac{1}{p}}{p} f_2
\]

Of course, \( f_1 \), the effort during forgetting, is not a part of the effort to be minimized because \( f_1 \) is zero and the forgetting time \( w_2 \) can be used effectively for other examinations, doing laboratory work, and the like. Now \( E(f) \) must be expressed as a function of \( t_{\text{max}} \). With that formula the computer can vary \( t_{\text{max}} \) until \( t_{\text{max}} \) is at its maximum knowledge of a student, who may follow the optimal strategy but who fails three times by chance, is given in Figure 1. The values of the parameters are: \( c/x = 0.1 \), \( m/i = 1.44 \), \( w_1 = 10 \), and \( t_{\text{max}} = 0.6 \).

The index 0 indicates the first learning period, the index 1 the first period of forgetting, and the index 2 the second learning period. Every time the same top is optimal and in principle the situation can be repeated infinitely.

Because forgetting increases with knowledge (as shown by the optimal strategy for the student is to study maximally \( g = 1 \) or not at all \( g = 0 \). Thus, though in the model \( g \) can have any value (below an upper limit), we can restrict ourselves to the case of \( g = 1 \) or \( g = 0 \).

The \( f_0 \) in (9) can be found from (4) with \( g = 1 \):

\[
f_0 = \frac{w_0}{1} \ln \left( 1 - \frac{t_{\text{max}}}{c m} \right)
\]

The calculation of \( f_0 \) given \( t_{\text{max}} \) is a bit more complicated. It may be convenient to accept the minimal knowledge at the moment in which the student begins to study again as the independent variable, in which case \( w_2 \) can be expressed with (3), \( t_b \) becoming \( t_{\text{min}} \) and \( t_0 \) becoming \( t_{\text{max}} \).
Likewise \( w_1 \) can be expressed with (5), with \( t_0 \) becoming \( t_{\text{max}} \) and \( t_c \) becoming \( t_{\text{min}} \):

\[
\frac{w_1}{m} = \ln \frac{t_{\text{max}}}{t_{\text{min}}}
\]  

(11)

\[
\frac{w_2}{m} = \ln \frac{cm}{xi} - t_{\text{min}}
\]  

(12)

With these two equations and

\[ w_1 + w_2 = w_t \]  

(13)

\( t_{\text{min}} \) can be solved:

\[
t_{\text{min}} = \frac{e^{m/ni}(cm/_{\text{max}}xi - 1)}{1 + e^{w1/mi}(cm/_{\text{max}}xi - 1)}
\]  

(14)

and with \( t_{\text{min}} \) and (12), \( w_2 \) can be found. It is realistic to equate \( f_2 \) in (8) not exactly with \( w_2 \) (with \( g = 1 \)) but to make it a bit larger:

\[ f_2 = w_2 + w_3 \]  

(15)

in which \( w_3 \) is formally equal to the duration of the examination itself. In reality, however, the examination takes more time for the student, in connection with the tension before and the relaxation afterwards, for example a full working day \( (w_3 = 0.2) \). This \( w_3 \) may be used as a personal parameter that expresses the aversion to fail.

In the constructed computer program \( t_{\text{max}} \) is systematically varied until \( t_{\text{opt}} \) is approached sufficiently accurately. Each time \( E(f) \) is calculated using the equations (1), (7), (10), and (12) to (15).

4. VARIANTS OF THE MODEL

Very little is known about learning and forgetting curves in a real educational setting. Curves other than those described above are also defensible. Indeed, the author has also used other models. In a first variant, learning took place with an exponential function with \( t_o = 1 \) as asymptote. This variant looks more like known mathematical learning theories but has the disadvantage that learning does not (as in reality) change gradually into forgetting when engagement decreases, but that the student can only be in two states: learning or forgetting [27]. In the model above only the values \( g = 1 \) and \( g = 0 \) are used, but other values are at least possible.

In a second variant an ability dimension going from minus infinity to plus infinity is used, as in test theory. Here, too, the increase of ability per time unit is proportional to \( c/x \), learning velocity and forgetting velocity are additive, and forgetting is proportional to \( 1/m \).

The model described above (linear learning superposed on exponential forgetting) seems to be the simplest possible, and thus the most acceptable, so long as little is known about the real curves. The linearity of learning (if forgetting is ignored) results from the assumption that, if the subject matter can be divided into parts that take equal times, most teachers will include equal numbers of items in every part of the examination.

In the following sections some examples will be given of possible applications. Though the parameter values are realistic, the examples are intended only to be illustrations, 'to stir one's mind' as Cronbach and Gleser [1] called it, to generate hypotheses, not to test them, and to show possibilities of solutions.

5. A FIRST EXAMPLE OF A POSSIBLE APPLICATION

A certain subject is examined twice a year \( (w_3 = 60) \) with a test of thirty-four-choice items \( (n_0 = 0.5) \), of which seventeen must be answered correctly. The parameter \( c/x \) can be estimated as 0.125 and \( m/1 \) as 66.7; \( w_3 = 0.2 \). The student union wants to increase the number of examinations to four a year. Calculations with the model show that this increase, without other changes, will lower the optimal strategy, \( t_{\text{opt}} \), of the student. The staff, however, wants to maintain this optimal knowledge level. Thus, if \( w_4 \) is lowered, the minimum passing score \( y \) must be increased. But this results in an increase of the expected total effort \( E(f) \). This last effect may be met by test lengthening.

The computer program constructed for this problem first calculates \( t_{\text{opt}} \) for the original situation of course after systematic variation of \( t_{\text{max}} \) and \( n_0 \), etc., until \( E(f) \) is minimized. This will be called the demand \( D \). Next \( w_4 \) is increased (in the example to 15) and the lowered \( t_{\text{opt}} \) calculated. If the \( y \) is increased again and again until \( t_{\text{opt}} \) exceeds \( D \). The procedure is repeated with a test one item longer, and again until \( E(f) \) is decreased below the original level.

It appears that the examination frequency can be enhanced from two to four times a year without decrease of the demanded knowledge level and without increase of the expected effort. So a situation satisfying to both parties can be created (if lengthening of the test meets no objections).

In the example the thirty-item test must be lengthened to 34 items with a minimum sufficient score of twenty. The demand \( D \) (0.571) then is just passed (0.573) and the expected effort is decreased from 3.29 to 3.05 study weeks.

To explore the influence of changes of parameter values the calculations were repeated with one parameter changed each time. It appears that if the subject matter is made five times as long \( (c/x = 0.025) \) the expected effort will be more than five times as large; the optimal knowledge level decreases clearly, as does the probability of success, but the calculated \( y \) remains constant. The same phenomenon can be seen if the initial knowledge, \( t_0 \), is varied. This suggests that the model may be used in these situations, for administrative purposes even if some parameter values are scarcely known.

As expected, test lengthening to 80 items leads to enhancement of the optimal knowledge level from 0.49 to 0.63 and decrease of total.
6. COMPARING CONJUNCTIVE AND COMPENSATORY COMBINATION OF TESTS

The following calculations were made in connection with the question: Given the subject matter, the tests, and the cutoff score, what is the most efficient way to combine tests: conjunctive or compensatory? Much is known about the efficiency of these methods, but only so far as the utility of the selected group is concerned. Not much is known about the effort of the students. Restricting ourselves to a concrete example: the effort necessary to pass (1) two tests of fifty items compensatorily combined (γ = 75), the student may do over again either test next time until the sum score is 75, will be compared with (2) two fifty-item examinations with γ = 38, conjunctively combined; (3) one test of fifty items (γ = 38, restriction of administration time); and (4) one test of 100 items (γ = 75, no objections to a long test).

The following parameter values will be assumed: c′ = 0.25 for one half of the total subject matter, thus 0.25 in Cases 1 and 2, 0.135 in Cases 3 and 4; m = 40, w = 8, g is 1 or 0, tp = 0.6 (two-choice items), w = 0.2. Cases 2, 3 and 4 can be solved with the program of Section 5. In Case 2 the effort to pass two tests conjunctively combined is twice the effort to pass one test. It appears that the method with one test of 100 items yields the least effort for the students (f = 3.70), one test with fifty items comes next, while the conjunctive method costs the greatest effort (3.63 weeks work). The sequence remains the same when the cutoff score is raised, but is reversed in the (unrealistic) case in which w is set to zero.

The significant part of this investigation, however, is the calculation of the effort when tests are combined compensatorily. This problem is much more complicated and we have to use a Monte-Carlo method with simulated scores. Given true score, the obtained score is assumed to be distributed binomially, and thus a score can be generated with the binomial distribution and a random number - p in equation (7), in which γ is now the score.

The (simulated) process begins with the first subject; estimates the optimal t, studies until this level is reached, takes the test (a score is generated). If the sum score is not yet 75 - and of course the first time it is not - he chooses the subject with his lowest score and estimates the optimal t. After a period of forgetting, in which his knowledge of both subject matters decreases, he studies until the subject is reached, takes the test, and so on. Groups of 100 persons are compared.

More details are given in [5]. Here we only summarize some results. It appears that the compensatory method costs the least effort. It can be said, however, that the mean sum score of the successful students with this method (76.85) is lower than the score if one test of 100 items is used (77.95). But if we raise the cutoff score of the compensatory method to 76, the mean score rises to 78.12, and in this case the effort is still smaller: 3.32 weeks against 3.70. So the calculations suggest that the compensatory method will be most efficient.

It must be recalled, however, that the two subject matters must have the same factorial composition. The model is still unidimensional, but a model with more subject matter dimensions may be more appropriate. Another restriction of the examples above is that all students are assumed to have the same parameter values. Variation of these values, however, will not fundamentally change the model.

7. ESTIMATION OF TRUE SCORE BY THE STUDENT

So far, it has been tacitly assumed that the student knows his true score exactly by means of preliminary examinations or published sets of items. In reality he can only estimate his true score. Random errors of estimation are important, not systematic errors. Let k be the number of items of the preliminary test, a random sample of the same domain as the examination, which consists of k2 items. The error variance of the examination score, given the true score t, is t(1 - t)/k2 according to the binomial error model. But this same model can be applied to the preliminary test. Here the true score is also approximately t, and thus the error variance is t(1 - t)/k1. Errors of the first and the second test are uncorrelated, thus

$$\text{total error variance} = t(1 - t)\left(\frac{1}{k_1} + \frac{1}{k_2}\right)$$

But this means that, as an approximation, the equations of the model can be used unchanged, provided that not the real number, k2, of the items of the examination is used in the equations, but a reduced number k, so that 1/k is the sum of 1/k1 and 1/k2. It follows that

$$k = \frac{k_1 k_2}{k_1 + k_2}$$

The shorter the preliminary test, the more the number of items of the examination must be reduced in the equations. If the examination and the preliminary examination are about equally long, or equally reliable, one may use \(\frac{k_2}{2}\) instead of k2.

A perspective

Other essential evaluations in higher education, such as research reports, laboratory work papers, essays, are not considered here. As far as examinations are concerned, however, a systems approach as outlined above may be used. The total subject matter can and must be
translated into questions, items, though multiple-choice items may not be necessary. Ideally one has to try to achieve a situation in which the student knows a certain percentage of the items of every subject matter in his later profession, but this ultimate criterion can seldom be used. One has to be content with a final examination as the intermediate criterion. The objectives are thus percentages of the item domains known on the last day the student is at the university.

Examinations given before the final examination, however, cannot be dispensed with. These examinations lead the learning process and force the student to reach mastery before going on. These advantages are known, of course, but until now examinations have been used only intuitively by instructors. A mathematical model in a systems approach can show how all the variables relate to each other and how an optimal solution can be outlined. This means that the requirements of the last examination can be met by all the students in the least time, by manipulating variables such as time interval between examinations, extent and treatment of the subject matter, and methods such as conjunctive and compensatory combination of scores.

Acknowledgement

I am much indebted to Fred Kerlinger for his many suggestions concerning the English text.

REFERENCES