Nederlands Tijdschrift voor de Psychologie 29, 431-446.

A MATHEMATICAL MODEL FOR THE OPTIMAL USE OF CRITERION REFERENCED TESTS

Robert F. van NAERSSEN
University of Amsterdam

A mathematical model is developed with which the total effort of the student can be minimized (thus the learning process optimized) for the following situation. The student must pass an examination or mastery test, but is allowed to do this again and again, with a fixed time between tests. He can estimate his true score by means of a preliminary examination; thus he is able to study until an optimal level is reached; that means, the expectation of his total effort is minimal, if the probability of failure is taken into account. It is assumed that true score is a normal-ogive - or logistic - function of ability. Forgetting is seen as a uniform velocity towards the left on the ability dimension. If 'engagement' is constant there is a uniform movement to the right. The velocity depends very simply on three personal parameters: 'engagement', 'capacity to learn the subject' and 'memory and on three subject matter parameters: 'length', 'difficulty,' and 'isolatedness'. It is shown hom, the parameters can be estimated empirically. A formula is developed with which the expectation of total effort is expressed as a function of these six parameters, true score, and probability of success. This probability is expressed as a function of true score, number of items, and cutting score. With this formula the optimal true score can be iteratively estimated. It is necessary to know this best tactic of the student before the learning and evaluation process can be made optimal.

1. Introduction

The model (or theory) which will be developed here concerns mainly educational situations in which well-defined examinations must be taken. The examination consists of a sample from a domain of items, of which the student must answer correctly a predetermined percentage, also known to him. By means of preliminary examinations and the publication of sets of items from the same domain the student is able to estimate his score, thus to what extent he is ready for the examination. If one is satisfied with norm referenced tests (relative norms) the optimal methods of Cronbach & Gleser (1965), and others, may be used. The situation with criterion referenced tests is different; here mathematical models for optimal use of tests are wanting or rudimentary (Carroll, in: Block 197 1).

A mathematical model would he relatively simple if preliminary examination and examination itself would be totally reliable. In this imaginary case the student could manage to succeed with every examination. One has to take into account that some students learn or forget faster than others, and that some books or courses are more difficult or more incoherent than others. A simple model for these problems will be developed in sections 2-9. In reality the student has, moreover, a risk of fail. We suppose that the student is allowed to take the examination again and again: there is no selection, only placement. This means that the student studies at his own rate. (This does not exclude group instruction, of course.) This situation is realized in continental universities. (There are some voices heard, which want to alter this situation in a more schoolish direction. On the other hand, in Anglo-Saxon countries, where the schoolish system dominates, there are people who propagate studying at an individual rate; especially supporters of programmed instruction and computer assisted instruction advocate an individual rate, but not only these.)

2. Learning curves and forgetting curves

The assumption of these curves is essential for the present theory. The learning curve gives the true score of the student on the item domain as a function of time, and the forgetting curve gives this true score in the time, after the examination, in which the student does not study.

There has been much laboratory research done on learning curves. One of the models used, for example, is the 'all or none' model. It is assumed that the student reads the subject matter which consists of items to be learned, say, n times. Each item has each time a probability c to be picked tip from memory. Once it is in memory it stays there (until the examination). With these assumptions the probability of the item being picked tip after n readings can be estimated and this is (practically) also the mean number of items the student knows after n readings, this means his relative score. This learning curve turns its concave side to the time-axis. The forgetting curve turns its convex side to this axis. The assumption here is that every item in memory has a probability c' to disappear. Learning and forgetting curves are both exponential functions, and each other's reflected image. With these curves some calculations for the optimal use of criterion referenced tests were given in e.g. Van Naerssen 1971.

Though impressive learning curve research has been done in the laboratory, this cannot be said about learning and forgetting in the educational setting. This research is just starting, and is as yet in the phase of testing hypotheses, not of describing curves. The convex and concave curves of the laboratory cannot be generalized into the educational situation. A weak point of the exponential curves, for example, is that the phenomenon of over-learning is not described. If a subject matter is learned over and over again, forgetting decreases. In the all-or-none model it makes little difference for remembering, if the subject matter is learned until 98% or until 99% is known, though learning until 99%, costs much more time.

Here a presumably better model is used, adopted from score theory. The true score is a (normal) ogive function of the 'underlying ability'. This 'test characteristic curve' is the sum of the 'item characteristic curves', that describe the probability of answering the item correctly as a function of the ability.

During the study phase the student moves on the ability dimension from left to right. This dimension is just what the examination measures. If the student ceases to study he moves to the left on the ability dimension.

Because exact empirical material is lacking, the simplest assumptions are made, namely that the student is moving with a uniform velocity, to the right or to the left. This means that learning curve and forgetting curve both are supposed to he ogives, though the former is (much) steeper. The difference with the exponential function concerns the beginning of the curve. One could imagine that, first, an orientation is needed during which the score on the test hardly increases. In the phase of forgetting there is at first a period in which the knowledge is still coherent. As this coherence decreases the knowledge falls faster into decay, until the decrement of the true score has to slow down, just as in the forgetting curves of the laboratory. The more the student (over)learns the subject matter, the more to the right he comes on the ability dimension and the longer it lasts before he has forgotten all. This is contrary to the situation in the other model, of the continually concave learning curve and the continually convex forgetting curve.

Forgetting goes on also during learning. It is af [sic] if the student is in a row boat on a river -- the Lethe -- of which the water flows slowly to the left. Without effort he floats on the stream towards ignorance, but during study he rows against the stream. The larger his effort or concentration the stronger his muscles (capacity to learn) and the lighter the boat (subject matter), the faster he reaches his object - a successful examination. And, of course, the stream can be fast or slow.

3. The formula for learning and forgetting

In accordance with the rule that one has to begin with the assumption of the simplest relations, it is postulated that the velocity of the student with respect to the stream -- thus the ability increment per unit of time equals the product or quotient of four magnitudes:

(1) the *enGagement* G, the measure in which the student concentrates on the subject matter;

(2) the Capacity C to learn the specific subject matter;

(3) the Length L of the subject matter, and

(4) the Difficulty D of this subject matter. The velocity of forgetting is simply the quotient of a subject matter parameter and a personal parameter:

(1) the Isolatedness I, the measure ill which the subject matter is incoherent and not integrated in a larger part of knowledge, and

(2) the Memory M, the measure in which the student is able to remember what he has learned.

Of course, one could speak about the coherence of the subject matter and the forgetfulness of the person, but for clarity the direction of the six 'factors' is chosen in such a way that the larger the three personal parameters - G, C and M - and the smaller the subject matter parameters L, D and I, the faster the subject matter is learned.

The change of the ability is written as ΔA, and time as W (from 'Weeks' and not as T from 'Time', because T is reserved for 'True score'). The formula for the ability increment per time, or the total 'Velocity' thus becomes:

$$V = \frac{\Delta A}{W} = \frac{G\,C}{L\,D} - \frac{I}{M} \qquad (1)$$

The handiest unit of time in an examination theory is probably the week. Months are not of' equal length, tile year is much too long. On the other side, the day is too short; here it is not considered that tile student does not study for one or two days of the week.

As for the ability, this magnitude is measured on an interval of which the values go from minus infinity to plus infinity. As is usual in test theory, it is assumed that true score is such an ogive shaped function of the ability that, as ability increases from minus infinity to plus infinity, true score increases from a basic value T_0 until the maximum value 1. Thus, true scores are expressed in relative units, which means that they are divided by the number of items k.

For later calculations it appears better to introduce a 'corrected' true score T_c, a linear function of T, but having extreme values 0 and 1:

$$T_c = \frac{T-T_0}{1-T_0} \tag{2}$$

One could imagine T_c as the proportion of items 'known', and assume that $T_0 = 1/a$, in which a is the number of alternatives of the test items. But the use of a basic value T_0 in the formulae is more flexible than' the use of the number of alternatives. In reality TO may deviate from $1/a$; usually T_0 is larger. Thus it may be better to estimate T_0 empirically, for instance by administering the test to a group which has not yet begun to study the subject matter.

A normal ogive can be assumed between T_c and A. But it has some mathematical advantages to replace the normal ogive by the logistic curve, which cannot be distinguished from it by the eye (Lord & Novick 1968: ch. 17). The relations between T_c and A, and in connection with (2) between T and A are:

$$T_c = \frac{1}{1+e^{-1.7A}} \tag{3}$$

and

$$A = \frac{1}{1.7}\ \ln\frac{T_c}{1-T_c} = \frac{1}{1.7}\ \ln\frac{T-T_0}{1-T} \tag{4}$$

The number 1.7 may be deleted, but usually it is added to conform the logistic curve to the standard normal-ogive one. With 1.7 the unit of ability is easier to interpret. Just as between $p = 0.5$ and about $p = 0.84$ of the standard normal ogive one standard deviation exists, one can say that the ability is increased by one unit (a standard deviation) if the corrected true score is increased from 50% to 84% (or from 16% to 50%, etc.).

On an interval scale, the zero point can be chosen arbitrarily. Here the ability is zero when T_c is 0.5.

The true score T must be estimated from observed scores Y (see appendix). The 'velocity' of learning or forgetting is expressed as 'standard deviations per week'. (Another method to get an easily interpretable ability increment, is to delete 1.7 in (4) but replace ln by ^{2}1og; this method is not used in the following sections).

5. Latent ability

When students have to take an earlier administered test unexpectedly, they earn (very) low scores. But this does not mean that little knowledge remained. They know the subject matter again perhaps after one rereading of the book. One has to postulate a latent knowledge behind the measured manifest knowledge. The manifest knowledge soon becomes very small, and, moreover, unreliable to measure. Latent knowledge may be more interesting; it expresses what 'really' remained of the once-learned subject, after reactivation. It is assumed in the model that the latent ability decreases linearly with time in the forgetting phase and increases linearly in the learning phase (if engagement is constant). The manifest ability may be a more or less capricious function of time.

A consequence of this assumption is that the ability can be measured only at the beginning of learning, and on the 'tops' of the curve, i.e. during. the examinations. It will be shown how the parameters of the model may be estimated with these assumptions.

6. Estimation of personal parameters

The units of four of the six parameters can be chosen arbitrarily. It appears to be easiest to choose the units of the three subject matter parameters and the unit of engagement. E.g. one chooses a 'typical' examination. The length (number of pages), the difficulty and the isolatednesss of this chosen subject matter define the units of length, difficulty and isolatedness.

As for the engagement, this can be expressed as a proportion of the customary maximal engagement of students in a week. E.g. 40 hours of intensive study per week defines an engagement of 100%. This engagementt is halved when the student divides his effort equally between two examinations, or his working time equally between examination and recreation. Of course, it also decreases when time is not optimally used. If one wants to use the model in short learning experiments, then one must take into account that the engagement may be 420% for some hours (a week has not 40 hours but 168).

In principle, the engagement score has to he obtained from the student, though a time-scheme and observations may be of some use. It must be clear from (1) that the larger the given engagement, the smaller the calculated capacity. The engagement as a separate factor is important when the effort is divided between several subjects (section 9).

Taking into account that the ability can only be measured in the beginning and at the 'tops' (during examinations), one can measure capacity and memory as follows by means of the defining subject matter: let the variables during the first study period be marked by the index 1, during the second, forgetting phase (after the failure of the examination) by index 2 and those of the second study period by index 3. Odd indices indicate study phases, even indices forgetting phases. Let A_0 be the ability at the beginning, A_1 the ability after the first (learning) period, etc. Then it follows from (1) and the fact that $L = D = I = 1$:

$$A_1 - A_0 = G_1 C W_1 - \frac{1}{M} W_1 \tag{5}$$

$$A_3 - A_1 = G_3 C W_3 - \frac{1}{M} (W_2 + W_3) . \tag{6}$$

These are two equations with the unknowns, C and M: linear equations if $1/M$ is regarded as the unknown. A_0, A_1, and A_3 can be estimated from scores; W_1, W_2, and W_3 can be measured or asked, just as the proportions G_1 and G_3. Thus, C and M of every student can be calculated.

In the past, intelligence was defined as the capacity to learn, but it will be clear that capacity and memory in this model are operationally defined variables, which have nothing to do with the problem of nature and nurture. In any case, C and M will increase with relevant education and later on decrease with age. The learning velocity of a student of statistics and the case with which he remembers formulae obviously increases with his former mathematical education. C and M are not constant, but they may be used as constants within the short time for which the model is designed: between beginning with the subject and passing the examination.

7. Estimation of subject matter parameters

When the values of the personal parameters are known, the parameters of other subject matters - not of the defining one - can be estimated. As for the length L, this parameter can easily be estimated in proportion to the defining subject matter; one may use the number of pages, of lessons, of formulae, etc., but, of course, once chosen a method must be maintained. The length L is the subject matter parameter which is the analogue of the personal parameter G. Likewise, the difficulty D is the analogue of capacity C and the isolatedness I the analogue of the memory M.

When the same student learns another subject matter, one can write down two equations with two unknowns again, comparable to (5) and (6), but now with the unknowns D and I, and known constants C and M.

Obviously, the same values of D and I must be found with another student. This is the first test of the model. Of course, it will be impossible to find exactly the same values with each student, but if these values vary too much the model must be adapted, i.e. made more complicated.

When L's, D's and I's of different subjects (examinations) are calculated, then. of course, these must produce the same personal parameters of the same students. This is the next test of the model.

Difficulty and isolatedness, too, are defined in terms of the velocity of learning and forgetting. The difficulty of a subject matter is in principie nothing more than the number of weeks it takes a certain person to learn a certain number of pages up to a certain level (proportion score), if forgetting during learning may be neglected. The isolatedness is nothing more than the velocity with which the subject matter is forgotten. It may be large when nonsense syllables must be learned, and small if tile subject matter is embedded in well-integrated knowledge.

The subject matter parameters are changing slowly in time. The course may be shortened or lengthened, but it is also possible to decrease D by publishing a syllabus, or by introducing working groups. The isolatedness may be decreased by showing relations between new concepts and old ones, by introducing open-book-examinations, etc. It is only assumed that L, D and I may be considered constant within a few months and without modification of the course.

It has been shown how, in principle, the six parameters may be estimated. Reality may be more complicated. Not exactly the same values of the personal parameters will be found, if one uses different subject matters, and not exactly the same values of subject parameters if one uses different students; one has to calculate means, minimal mean squares, etc. Here it was only the purpose to show that the theory can be tested empirically, and how this can be done in principle. In section, 9 the model will be tested otherwise.

Each theory or model is a simplification. One simplification here is that the subject matter is regarded as homogeneous. In reality one could speak about learning items, which each have their own learning curve and forgetting curve, analogous to the item characteristic curves of score theory. If the item characteristics are normal ogives, for example, then the test characteristic curve is only approximately normal. The probability of a right answer to some learning items will soon be decreased to the basic value, other items are remembered a whole life span, etc. Nevertheless: the model may be useful within the short interval between beginning a course and passing the test.

8. Extent of subject matter and total effort

Often it will be unnecessary to distinguish between length and difficulty. In these cases their product may be indicated by the word 'extent' and the letter X.

A second complex concept is very important in this model: the (total) effort F. When engagement is constant this effort is the product of engagement and time. More generally:

$$F = \int G \, \mathrm{d} W \quad \text{or:} \quad F = \sum_i G_i W_i \tag{7}$$

The concept 'effort' is central in this model, because it is assumed that if the educational process be optimal, both student and teacher (staff') have to minimize the total effort (of the students, given the extent of the subject matter and the required level of competence).

Roughly speaking the effort F is proportional to the extent X of the subject matter, but this is only true insofar as the term $1/M$ may he ignored. When $1/M$ approaches GC/LD (see formula 1) the effort becomes infinite.

9. Testing the model: split of subject matter

Sometimes one hears about student actions in favour of splitting large examinations. They do not protest against the extent X of the subject matter, nor against the demanded level T (proportion of items which must be answered correctly) but require that the subject matter be spread over more -- say j equivalent - parts, which may be learned and passed in succession. Apparently the total effort is decreased. Does the model also lead to the same conclusion?

By resolving W in (1) and, multiplying this by (the constant) G, one gets a formula for the total effort F:

$$F = \frac{\Delta A}{\frac{C}{X} - \frac{I}{GM}} \qquad (8)$$

After the subject matter is divided into j equivalent parts, Δ A remains the same (because T is constant), and C, I, M, and G do not change either. The effort needed for each part is given by (8) on the understanding that X must be divided by j. The total effort is j times as large; thus, the total effort TP after division into j parts, becomes:

$$F_j = \frac{\Delta A}{\frac{C}{X} - \frac{I}{jGM}} \qquad (9)$$

If (8) and (9) are compared, it is seen that, indeed, F is decreased. The intluence of the division is largest when the second term in the denominator is a relatively large part of the first term; thus, if capacity, memory and engagement are relatively small and the extent and the isolatedness of the subject matter are large. In these cases students should strive towards division, in accordance with common sense.

It is also clear that without division a subject may be so heavy (*X* and *I* large) that a weak student needs an infinitely long time to pass the examination. Thus, in this model it is not assumed that 'everybody can pass every examination if lie gets time enough'. If the extent *X* of the subject matter is increased the weak students will fail first, but at last every student can be made to fail, because forgetting goes on as fast as learning and mastery is never approached.

By making the extent large enough every examination can be made selective; that means that only students with high capacity and memory will pass. On the other hand every weak student can be made to pass a difficult subject matter, provided that one tests in small parts. This strategy, of course, has also an obvious disadvantage: the subject matter may be wholly forgotten by the time the student graduates, while usually the objective is that the students dispose of the knowledge especially after that day.

10. Cutting score and probability of success

Now the examination *Y* with its fallible score *Y* enters into the picture. Only criterion referenced tests are considered, with a cutting score Y_s which is laid down beforehand, and known to the student. Scores Y_s or higher are considered 'mastery' or 'satisfactory'. If the student fails he is not removed, but he gets the opportunity, after some time, to take the test again and again (otherwise one could speak about selection).

The probability of success can be expressed as a function of three variables: cutting score *Y*, number of items *k* of the test, and true score *T* (see section 4). *Y*, and *k* must be manipulated by the staff if they want to optimize the educational process. The true score *T* to be aimed at is chosen by the student by means of preliminary tests or published sets of items. The chosen *T* is the 'strategy' of the studen *t*.

The binomial model (Lord & Novick 1968: ch. 23) will be used to express the probability *p*, of a student with true score *T*. It is assumed that the test items have (nearly) equal probability of success. The probability of a successful examination is the sum of the probabilities of having scores at least as large as Y_s :

$$p_s = \sum_{y=Y_s}^{k} \binom{k}{y} T^y (1-T)^{k-y} \tag{10}$$

(An alternative which appears less realistic is the model of knowing or blind guessing (Lord & Novick 1968: ch. 14). If one wants to apply this model the probability of success can be calculated as follows: it is assumed that $T_0 = 1/a$ and that the student knows $T_c k$ items and answers them correctly; he guesses blindly at the other C items: $k' = (1-T_c)k$. Each time the probability of success is $1/a$. If he knows at least as much as the cutting score he succeeds with certainty:

If $T_o k \geq Y_s$ then $p_s = 1$

If $T_c k \leq Y_s$ then $p_s = \sum_{y=Y_s-T_c k}^{k} \binom{k'}{y} \left(\frac{1}{a}\right)^y \left(1-\frac{1}{a}\right)^{k'-y}$ (11)

Thus the model of knowing or blind guessing.)

11. Minimizing the total effort

If norm-referenced tests are used the result may come as a surprise to the student, because he cannot know how hard the other students have learned and tried; the cutting score depends on their effort. This means that if norm-referenced tests are used, the student cannot design an optimal strategy, he is only able to guess to what extent he has to study the subject. But if the student cannot choose an optimal strategy, neither is the staff able to design an optimal educational system. All measures are mere guesses. If, on the other hand the cutting score is established in advance, and the student is able to estimate his true score, then he can choose an optimal level of knowledge A_{opt} with which the expectation of his total effort $E(F)$ is minimal.

The higher his chosen level A, the larger the effort necessary to reach this level, but the smaller his probability of failure. If he fails, his ability decreases. More effort is needed to reach the optimal level again. If his chosen level is too low he fails too many times. There must be somewhere an optimal level A_{opt} and, of course (formula 4) an optimal true score T_{opt}. If the student chooses the optimal true score T_{opt} he nevertheless may fail. The second time he has to reach the same optimal level, because every time the situation is the same; every time he has the same probability to succeed, etc.

The expectation of the total (optimal) effort $E(F)$ may be divided into three parts: the effort F_0 needed to increase F_c from some value above F_0 - due to earlier education - to the value 0.5; an effort F_1 needed to bring F_c to the optimal value T_{opt}, and the expectation of the effort F_2. F_2 is the effort needed to reach the optimal level again (after failure). If the probability of failure is called q_s (= 1-p_s) the expectation of F_2 is

$$E(F_2) = q_s F_2 + q_s^2 F_2 + q_s^3 F_2 + \ldots = \frac{1-p_s}{p_s} F_2 \,. \tag{12}$$

The effort F_2, needed to bring the ability level back to the same value is independent of study methods in this model; the student may do other things first and postpone the work, or he may study every day for some time. This can be proved as follows. If the level returns to the same value, then, with summation over time periods,

$$\sum_i (\Delta A)_i = \sum_i W_i \left(\frac{G_i C}{X} - \frac{I}{M} \right) = 0 \tag{13}$$

Because $\sum W_i = W$, the number of weeks between successive tests:

$$F_2 = \sum_i W_i G_i = \frac{X\,I}{C\,M} \; W \tag{14}$$

The effort needed to keep the ability (knowledge) on a constant level appears to be proportional to the time between successive examinations, and to extent and isolatedness of the subject matter; it is inversely proportional to capacity and memory.

Thus, for a certain person in a certain test situation F_2 is constant and can be estimated by (14). The expectation of total effort can be expressed as a function of the optimal true score T_{opt}. The second term follows from (4) and (8):

$$E(F) = F_0 + \frac{\frac{1}{1.7} \ln \frac{T_{opt} - T_0}{1 - T_{opt}}}{\frac{C}{X} - \frac{I}{GM}} + \frac{1-p_s}{p_s} \; \frac{X\,I}{CM} \, W \tag{15}$$

In this equation p_s can be calculated with (10), in which T is replaced by T_{opt}, from T_{opt}, the number of items k, and the cuttingscore Y_s.

However, the real unknown is just T_{opt}. This value can be approximated iteratively as that value between T_0, and 1, with which one gets the lowest value of $E(F)$ (this expectation must be minimized). The constant F_0 is of no interest when T_{opt} is calculated.

When T_{opt} is known, the optimal strategy of the student is known. If this is explained to the student, it may be expected that this strategy is followed. Then the probability of success p, is known, likewise the expectation of the total effort $E(F)$. By means of computer simulations variables can be manipulated such as the time W between examinations, the number of items k, the cutting score Y, the number of tests in which the subject matter must be divided; several methods of combining scores can be tested (compensatory vs conjunctive, etc.). In short: an optimal system of examinations can be designed.

Applications of the model, in the form of simulations, will be postponed till a following paper.

Appendix: True score estimated by the student

So far, it was tacitly assumed that the student knows his true score by means of preliminary examinations or published sets of items. In this way he is able to study until his optimal true score is reached. However, in reality he can only estimate his true score. How does this fact influence the model?

One has to distinguish between systematic and random errors. The observed score on the preliminary test may deviate systematically from the true score. One well-known method is to estimate the true deviation score as the observed deviation score times the reliability of the (preliminary) test. Other methods may be used. But, if it can be assumed that every student follows the optimal tactic, score variance is zero and the estimated true score is just the observed score of the preliminary test.

The random errors are much more important in the model. They can be handled as follows. Let k_1 be the number of the preliminary items - a random sample of the same domain as the examination, which consists of k_2 items. The error variance of the examination, given a true score T, is $T(1\text{-}T)/k_2$ according to the binomial 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 $T(1\text{-}T)/k_1$. Errors of' first and second test are uncorrelated, thus

total error variance =

$$T(1-T)\left(\frac{1}{k_1}+\frac{1}{k_2}\right) \tag{16}$$

But this means that the formulae of the model can he used unchanged, provided that not the real number k_2 of the items of the examination is used in the formulae, but a reduced number k so that $1/k$ is the sum of $1/k_1$ and $1/k_2$. It follows that

$$k=\frac{k_1 k_2}{k_1+k_2} \tag{17}$$

The shorter the preliminary test, the more the number of items of the examination must be reduced in the formulae. It turns out that the length of the preliminary test, with which the student estimates his true score, is just as important as the length of the examination itself.

References

Cronbach L.J. and G.C. Gleser, 1965. Psychological tests and personal decisions (2nd. ed.). Urbane, Ill.: Univ. of Illinois Press.

Block, J.11. (ed.), 1971. Mastery learning. New York: Holt, Rinehart & Winston.

Lord, F.M. and M. Novick, 1968. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.

Naerssen, R.F. van, 1971. Een model voor tentamens (vervolg). Ned. Tijdschr. Psychol. (D.P.O.) 26, 551-559.