DOCUMENT RESUME

ED 052 255	TM 000 661
AUTHOR	Moore, William E.
TITLE	Stochastic Processes as True-Score Models for Highly Speeded Mental Tests.
INSTITUTION	Educational Testing Service, Berkeley, Calif.
REPORT NO	RB-70-66
PUB DATE	Nov 70
NOTE	35p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Arithmetic, Goodness of Fit, Grade 10, *Mathematical Models, *Mental Tests, *Probability, Scores, Statistical Analysis, Statistics, Testing, *Timed Tests, Transformations (Mathematics), *True Scores
IDENTIFIERS	Erlang Process, *Kit of Reference Tests for Cognitive Factors, Poisson Process

ABSTRACT

The previous theoretical development of the Poisson process as a strong model for the true-score theory of mental tests is discussed, and additional theoretical properties of the model from the standpoint of individual examinees are developed. The paper introduces the Erlang process as a family of test theory models and shows in the context of mental testing, that the Poisson process is a particular case of the Erlang process. Probability density functions mathematically define the parameters and lead to semantic interpretations of parameters in terms of tests and examinee characteristics. Experimental research gives the fit of observations to theoretically predicted functional forms for individual examinees. In particular, experimental measurements determine observed-score distributions for individuals and give estimates of their true scores. The models apply to homogeneous, itemized tests of pure speed. The models were tested with tenth grade boys who responded to six highly speeded tests from the Kit of Reference Tests for Cognitive Factors. Three tests measured speed of clerical or perceptual skill, and three measured speed of computational skill in arithmetic. The Poisson process was applied with varying degrees of success to three of the six highly speeded tests while the Erlang process applied to one test. For two tests (Maze Tracing and Addition) the stochastic models had no application. (Author/CK)

ERIC

ED052255

RESEARCH BULLETIN

RB-70-66

STOCHASTIC PROCESSES AS TRUE-SCORE MODELS FOR HIGHLY SPEEDED MENTAL TESTS

William E. Moore
Harvard University, Cambridge

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

TM 000 661

Educational Testing Service
Princeton, New Jersey
November 1970

ED052255

STOCHASTIC PROCESSES AS TRUE-SCORE MODELS FOR HIGHLY SPEEDED MENTAL TESTS[1]

William E. Moore

Harvard University, Cambridge[2]

This paper discusses previous theoretical development of the Poisson process as a strong model for the true-score theory of mental tests. It develops additional theoretical properties of the model from the standpoint of individual examinees. The paper introduces the Erlang process as a family of test theory models and shows, in the context of mental testing, that the Poisson process is a particular case of the Erlang process. Probability density functions define the parameters mathematically and lead to semantic interpretations of parameters in terms of tests and examinee characteristics. Experimental research gives the fit of observations to theoretically predicted functional forms for individual examinees. In particular, experimental measurements determine observed-score distributions for individuals and give estimates of their true scores. The models apply to homogeneous, itemized tests of pure speed.

The classical model for the true-score theory of mental tests (Lord and Novick, 1968; Novick, 1966) relates observed, true and error scores for an examinee, i , on a given test as

---

[1]The initial impetus for this study developed while I was a Summer Fellow at Educational Testing Service in 1967. Professor William Meredith of the University of California at Berkeley, who was then spending a year at ETS, was my adviser in the Summer Program. Professor Meredith introduced me to the Poisson process as a test theory model and I gratefully acknowledge the conversation in which he suggested the Erlang process as a source of further test theory models. Professor Frederick Mosteller was my adviser at Harvard University and I thank him for invaluable advice and untiring interest during this research project.

[2]Now, Chief of the Centre for Research in Measurement and Evaluation, Department of Education, New South Wales, Australia.

(1) $$X_i = T_i + E_i \quad .$$

The observed test score, X_i, and the error test score, E_i, are random variables and the true score, T_i, is a constant defined by

(2) $$\mathcal{E}X_i = T_i$$

where $\mathcal{E}$ is the operator expectation. To estimate expected values we require repeated measures and unless these are available the classical model has nothing to say in an operative sense about the true scores of individuals.

The classical true-score model is the simplest case of weak theory, meaning the absence of specific assumptions concerning functional forms for observed scores and error scores. The stochastic models are examples of strong theory which specifies functional forms for random variables. We pay a price for the increased mathematical richness of strong models. Whereas the weak classical model is useful for any rigorously constructed mental test, the stochastic models restrict themselves to homogeneous, itemized tests of pure speed. In compensation for this restriction we are able to estimate the true scores of individual examinees from a single administration of a test.

Pure speed implies that examinees are under time pressure and, in fact, none should be able to complete all items within the imposed time limit. It also implies that items be of such a nature that all examinees would obtain correct answers if given sufficient time. Itemized means that responses are scored as being either right or wrong and homogeneous means that items are of equal difficulty--that is, the probability of an examinee answering correctly is the same for all items.

The Poisson Process

The Poisson is a stochastic process in which equiprobable events occur in continuous time. Visualizing each event as a point on the time axis and considering the random placement of points, the Poisson process postulates exponentially distributed time gaps (or latencies) separating points and postulates that the number of points recorded in a fixed time interval is the realization of a Poisson random variable.

Following Rasch's (1960) initial explorations, Meredith (1968) developed the Poisson process as a true-score model for speeded mental tests. The probability density function for the random variable representing observed score for examinee i on a test of duration v units in time is a form of the Poisson distribution. Letting Y_{iv} represent the observed-score random variable for examinee i, we write

$$(3) \qquad \Pr(Y_{iv} = y \mid i) = \frac{e^{-\Omega_{iv}} \Omega_{iv}^{y}}{y!} \quad .$$

Equation (3) is conditional on i to emphasize reference to a fixed person. From the first moments of equation (3),

$$(4) \qquad \mathcal{E}Y_{iv} = \Omega_{iv} \quad .$$

Thus, by definition from the classical model, Ω_{iv} is the constant representing the true score for examinee i. In terms of the Poisson process, Ω_{iv} is the mean value function and is related to the intensity function (or, the infinitesimal response rate) $\Lambda_i(t)$ by

(5) $$\Omega_{iv} = \int_{o}^{v} \Lambda_i(t)dt$$

(see Karlin, 1966; Meredith, 1968; Parzen, 1962).

Meredith postulates that the intensity function is the product of two independent components,

(6) $$\Lambda_i(t) = \lambda_i L(t)$$

with λ_i a positive constant for examinee i and $L(t)$ a continuous and positive function for $t > o$. The postulate asserts that the intensity function of different examinees has the same shape but different elevations. The Poisson process requires latencies to be exponentially distributed random variables. If t_j is the latency associated with the jth response for examinee i and if we write the transformation $L(t_j)$ as t^*_j, then the probability density function for latencies is

(7) $$f(t^*_j|i) = \lambda_i e^{-\lambda_i t^*} \quad .$$

Equation (7) completes the Rasch-Meredith formulation of the Poisson process as a true-score model for individual examinees. In conclusion we note that if an examinee's number of responses is the realization of a Poisson random variable, it is a necessary and sufficient characterization of a Poisson process that the latencies be exponentially distributed (Haight, 1967).

Transforming Time

To find a suitable transformation function of time we adopt the Bush-Mosteller procedure for analyzing the Weinstock data on timing rats in maze trials (Bush and Mosteller, 1955). Transferring their approach to the present situation we let the time for response j from examinee i be

(8) $$t_j = c_i + \tau_j$$

where c_i is a constant element of time particular to examinee i and τ_j is, as the Poisson process requires, an exponentially distributed random variable specified by equation (7). For mental testing, equation (8) implies that an examinee requires a certain minimum (constant) length of time to respond to any item and the time he takes beyond this limit is a random variable. As t_j is the manifest variable, equation (7) is more convenient in the form

(9) $$f(t_j|i) = \lambda_i e^{-\lambda_i \tau} \quad , \qquad j = 1,2,\ldots \ .$$

In speeded tests we assume examinees make their responses instantaneously so that the time required for k responses is the sum of k latencies. Considering an examinee's latencies for k items relates the sequence of latencies to the number of responses he records in fixed time. These latencies are

(10) $$t_K = \sum_{j=1}^{k} t_j \quad .$$

Moment generating functions show the k-fold convolution of t_j to be gamma distributed with density function

(11) $$g(t_K|i) = \frac{\lambda_i}{(k-1)!} (\lambda_i \tau_K)^{k-1} e^{-\lambda_i \tau_K}$$

where τ_K is

(12) $$\tau_K = t_K - kc_i \quad .$$

The number of responses in fixed transformed time, τ' , is k if and only if $\tau_K \leq \tau'$ and $\tau_{K+1} > \tau'$. Equivalently (Parzen, 1962, p. 132),

(13) $$\Pr(y_{i\tau'} = k|i) = \Pr(\tau_K \leq \tau'|i) - \Pr(\tau_{K+1} \leq \tau'|i) \quad .$$

We have

(14) $$\Pr(\tau_K > \tau'|i) = \int_{\tau'}^{\infty} \frac{\lambda_i}{(k-1)!} (\lambda_i \tau_K)^{k-1} e^{-\lambda_i \tau_K} d\tau_K$$

which on integration by parts gives

(15) $$\Pr(\tau_K > \tau'|i) = \sum_{m=o}^{k-1} \frac{1}{m!} (\lambda_i \tau')^m e^{-\lambda_i \tau'} \quad .$$

Simple manipulation of equation (15) and substitution into equation (13) give

(16) $$\Pr(Y_{i\tau'} = k|i) = \frac{1}{k!} (\lambda_i \tau')^k e^{-\lambda_i \tau'}$$

which is a form of the Poisson distribution.

From equations (3), (12) and (16) we get

(17) $$\Omega_{iv} = \lambda_i [\sum_{j=1}^{y} t_j - yc_i] \quad .$$

Since $\sum_{j=1}^{y} t_j$ is the length of the test in ordinary time, we may write more simply

(18) $$\Omega_{iv} = \lambda_i [v - yc_i] \quad .$$

The probability of an examinee's response to the jth item falling in an arbitrary interval $(a,b]$ of ordinary time is from equation (9),

(19) $$Pr(a \leq t_j < b | i) = \int_a^b \lambda_i e^{-\lambda_i (t-c_i)} dt = e^{-a*} - e^{-b*} \quad ,$$

where

(20) $$a* = \lambda_i (a - e_i) \quad \text{and} \quad b* = \lambda_i (b - c_i) \quad .$$

Transforming time by

(21) $$t* = \lambda_i (t - c_i)$$

enables us to construct a table of theoretical probabilities which applies to all examinees and leads to two important theoretical predictions. First, summing equation (21) over the y latencies given by examinee i in a test of length v units of ordinary time gives

$$(22) \qquad \sum_{j=1}^{y} t_j^* = \lambda_i[v - yc_i] \quad .$$

Equations (18) and (22) show that the mean value function of the Poisson process, Ω_{iv} , equals the length of the test in transformed units. From equation (4) we conclude that an examinee's true score is the length of the test for him in transformed time although the test is fixed to a common length in ordinary time for all examinees. Second, partitioning transformed time into y equal parts produces realizations of y Poisson random variables each of which has a mean value function of unity. Thus, we expect from theory each examinee to produce one response in each unit of transformed time.

The Erlang Process

The Erlang, like the Poisson, is a stochastic process in which equiprobable events occur in continuous time. The Erlang process requires, for its characterization, gamma distributed latencies and an Erlang distribution for the random variable representing number of events in fixed time. An exponential distribution of latencies is a foundation of the Poisson process, and we have noted in connection with equation (11) that the k -fold convolution of an exponential distribution is a gamma distribution. An interpretation of the Poisson model which is conceptually attractive to mental testing is that examinees are responding to items which require one step to solve. In pursuing this image we now develop the Erlang process as a family of stochastic models in which examinees respond to items requiring 2, 3 or more steps to complete.

Let us assume that examinees require n unobservable subresponses to complete each response. Let t_0 be the time required for a subresponse and assume, as we have for the Poisson process (equation 8), that t_0 is some constant h_i plus a random variable τ. That is

$$t_0 = h_i + \tau \quad . \tag{23}$$

As with the Poisson model, we write

$$f(t_0|i) = \lambda_i e^{-\lambda_i \tau} \quad , \qquad 0 = 1,2,\ldots,n \; . \tag{24}$$

The total latency on the jth response by examinee i is nh_i plus the sum of n random variables τ from equation (24). Letting the total latency for the jth response be L_{ij} gives

$$\begin{aligned} \phi(L_{ij}|i) &= \frac{\lambda_i}{(n-1)!} [\lambda_i(L_{ij} - nh_i)]^{n-1} e^{-\lambda_i(L_{ij}-nh_i)} && \text{for } L_{ij} \geq nh_i \\ &= 0 && \text{for } L_{ij} < nh_i \; . \end{aligned} \tag{25}$$

The latencies for the examinee's responses to k items are $\sum_{j=1}^{k} L_{ij}$, of which the random variable component is $(\sum_{j=1}^{k} L_{ij} - knh_i)$. By considering moment generating functions we find that the k-fold convolution of $\sum_{j=1}^{k} L_{ij}$ is the gamma function

(26) $$\psi(L_{iK}|i) = \frac{\lambda_i}{(nk-1)!}[\lambda_i(L_{iK} - knh_i)]^{nk-1} e^{-\lambda_i(L_{iK}-knh_i)} \quad \text{for } L_{iK} \geq knh_i$$

$$= 0 \quad \text{for } L_{iK} < knh_i$$

where

(27) $$L_{iK} = \sum_{j=1}^{k} L_{ij} \quad .$$

To derive the probability density function for the number of responses in fixed time we repeat the argument involving equations (12) through (16). Thus,

(28) $$\Pr\{[L_{iK} - knh_i] > \tau'|i\} = \int_{\tau'}^{\infty} \psi(L_{iK}|i)dL_{iK}$$

which on integration by parts gives

(29) $$\Pr\{[L_{i\kappa} - knh_i] > \tau'|i\} = \sum_{m=o}^{nk-1} \frac{1}{m!}(\lambda_i\tau')^m e^{-\lambda_i\tau'} \quad .$$

Similarly,

(30) $$\Pr\{[L_{i(\kappa+1)} - (k+1)nh_i] > \tau'|i\} = \sum_{m=o}^{(k+1)n-1} \frac{1}{m!}(\lambda_i\tau')^m e^{-\lambda_i\tau'}$$

which leads to

$$(31) \qquad \Pr(Y_{i\tau'} = k|i) = \sum_{m=nk}^{(k+1)n-1} \frac{1}{m!} (\lambda_i \tau')^m e^{-\lambda_i \tau'} \quad .$$

For a test of length v units of ordinary time the probability of examinee i giving y responses is

$$(32) \qquad \Pr(Y_{iv} = y|i) = \sum_{m=ny}^{(y+1)n-1} \frac{1}{m!} (\Omega^*_{iv})^m e^{-\Omega^*_{iv}} \quad ,$$

where

$$(33) \qquad \Omega^*_{iv} = \lambda_i [v - ynh_i] \quad ,$$

that is, $\Pr(Y_{iv} = y|i)$ is a summation of Poisson functions between the limits ny and $ny + (n - 1)$. As a probability density function it is known as the Erlang after the Danish statistician who first applied it to problems in receiving calls at telephone exchanges.

If an examinee is responding to a speeded test with homogeneous items each requiring n steps to complete, then the Erlang process predicts that equation (25) is the probability density function for his distribution of latencies and predicts that equation (32) is the probability density function for his distribution of number of responses in fixed time. The transformation function for time is the same as for the Poisson model (equation 21) with c_i being replaced by nh_i . Equation (33) shows that the parameter of the Erlang distribution, Ω^*_{iv} , equals the length of the test in transformed time.

The term n in equations (25), (30) and (33) is a parameter which did not appear in the Poisson model. We discuss parameters in the next section, but for present discussion we consider the special cases when n equals 1 or 2.

When $n = 1$ equation (25) becomes

$$\phi(L_{ij}|i) = \lambda_i e^{-\lambda_i(L_{ij}-h_i)} \quad . \tag{34}$$

Equation (34) is in the same exponential form as equation (7). For $n = 1$, equation (32) reduces to equation (3). Because the gamma function has the special property of collapsing to an exponential for $n = 1$, and because the Erlang function collapses to a Poisson function for $n = 1$, we acknowledge the Poisson process as being a special case of the Erlang process.

When $n = 2$ equation (25) becomes

$$\phi(L_{ij}|i) = \lambda_i[\lambda_i(L_{ij} - 2h_i)]e^{-\lambda_i(L_{ij}-2h_i)} \tag{35}$$

and equation (32) becomes

$$\Pr(Y_{iv} = y|i) = \frac{1}{2y!}(\Omega^*_{iv})^{2y}e^{-\Omega^*_{iv}} + \frac{1}{(2y+1)!}(\Omega^*_{iv})^{2y+1}e^{-\Omega^*_{iv}} \quad . \tag{36}$$

Equations (35) and (36) are the specification equations for the simplest Erlang model with gamma distributed latencies. As n takes higher integral values, $n = 3,4,\ldots$, we generate a family of Erlang models.

To deduce the theoretical relationship between number of responses and length of transformed time we consider first moments of equation (36). The expected value of Y_{iv} is (Parzen, 1962)

$$(37) \qquad \mathcal{E}(Y_{iv}) = \frac{\Omega^*_{iv}}{2} - \frac{1}{4} - \frac{1}{4} e^{-\Omega^*_{iv}} \quad .$$

Equation (33) establishes that Ω^*_{iv} equals the length of the test in time transformed by equation (21). For 10 units of transformed time equation (37) becomes

$$(38) \qquad \mathcal{E}(Y_{iv}) = \frac{10}{2} - \frac{1}{4} - \frac{1}{4} e^{-20} \quad .$$

As $\frac{1}{4} e^{-20} \sim 0$, this model predicts that an examinee will respond to 4.75 items in each ten units of transformed time.

We conclude that an examinee's true score is 0.475 of length of the test for him in transformed time although the test is fixed to a common length in ordinary time for all examinees. Further, we expect each examinee to respond to 0.475 items in each unit of transformed time.

The Parameters

The exponential distribution of latencies in the Poisson process is a special case of the gamma distribution of latencies in the Erlang process. We take advantage of this relationship to develop a single set of procedures applying to parameters of both stochastic processes. The common equation for latencies is

$$(39) \qquad \phi(L_{ij}|i) = \frac{\lambda_i}{(n-1)!}[\lambda_i(L_{ij} - c_i)]^{k-1} e^{-\lambda_i(L_{ij}-c_i)} \quad , \quad j = 1,2,\ldots \quad L_{ij} \geq c_i$$

$$= 0 \qquad\qquad L_{ij} < c_i \; .$$

The term L_{ij} is the latency measured in ordinary time for subject i on item j of a test in a particular trial. The terms n, λ_i and c_i are parameters.

The parameter n identifies a test with a member of the family of gamma distributions and is the test parameter. In these models the test parameter takes only positive integral values of one or greater. This restriction is imposed to give necessary mathematical tractability to equation (31). The structure of the models suggests that the gamma function, equation (39), is formed by the convolution of n exponential distributions having the same argument. A semantic interpretation of n which is conceptually attractive to test theory is that the Poisson model (in which $n = 1$) applies to tests in which each item requires a single step to solve. The simplest Erlang model has $n = 2$. This model applies to tests in which each item requires two identical steps to solve. For example, in subtracting one 2-digit number from a larger 2-digit number we could reasonably suppose that the subtraction of the units' digits is one step and the subtraction of the tens' digits is a second and equal step. Similarly, for $n = 3,4,\ldots$ we have Erlang models applying to tests with items that require $3,4,\ldots$ steps to complete.

From the axioms of the Poisson process (Meredith, 1968) we know that $\lambda_i dt$ is the probability of an examinee producing a response in an instant of time. The parameter λ_i is particular to examinee i on a given trial and is the rate parameter. The parameter c_i is the constant-time parameter and is the nonrandom (constant) component of a latency. It is particular to subject i for all responses on a given trial.

Bush and Mosteller (1955) devised methods for estimating n , λ_i and c_i which make use of percentage points of the cumulative frequency distribution of latencies. When estimates of n for different examinees give a common integral value we identify the Poisson model or the particular Erlang model which may apply to this test. Estimates of λ_i and c_i transform time into units particular to each examinee. The frequency distributions of transformed latencies are compared with the appropriate theoretical distribution based equation (39). The probability table has class intervals initially of 0.25 units and as probabilities decrease, the class intervals increase to 0.50 units, to 1.00 units and finally to 2.00 units.

Experimentation

The subjects were ten Grade 10 boys aged 15 to 16 years from a Massachusetts high school and they responded to six highly speeded tests from the Kit of Reference Tests for Cognitive Factors (French, Ekstrom and Price, 1963). Three of the tests measure speed of clerical or perceptual skill and three measure speed of computational skill in arithmetic. Each test is in two parallel forms of sufficient length, in some cases, to subdivide each form into two parts. The subjects completed up to four trials on each test and trials lasted from 1 1/2 to 3 minutes each.

Tests were administered to one subject at a time. During testing, each subject was tied in to an electronic system designed to compute latencies. At

Table 1

Expected Probabilities for the Distribution of

Latencies in Transformed Time

Transformed Time	Exponential Probabilities	Gamma (n = 2) Probabilities
$t^* = \hat{\lambda}_i(t - \hat{c}_i)$	$Pr = \int_a^b \lambda_i e^{-\lambda_i(t-c_i)} dt$	$Pr = \int_a^b \lambda_i^2 (t-c_i) e^{-\lambda_i(t-c_i)} dt$
0.00 - 0.25	0.221	0.0284
0.25 - 0.50	0.172	0.0611
0.50 - 0.75	0.135	0.0839
0.75 - 1.00	0.104	0.0923
1.00 - 1.25	0.081	0.0876
1.25 - 1.50	0.064	0.0883
1.50 - 1.75	0.049	0.0813
1.75 - 2.00	0.039	0.0693
2.00 - 2.50	0.053	0.1208
2.50 - 3.00	0.032	0.0873
3.00 - 3.50	0.020	0.0644
3.50 - 4.00	0.012	0.0437
4.00 - 5.00	0.0116	0.0512
5.00 - 6.00	0.0043	0.0229
6.00 - 8.00	0.0021	0.0144
8.00 -10.00	0.0003	0.0025

Note: The terms $\hat{\lambda}_i$ and $\hat{c}_i$ refer to estimated values of parameters λ_i and c_i.

the moment a subject recorded a response a print-counter recorded the elapsed time from the previous response correct to one-tenth of a second. The printed records of each trial told the exact length of the trial in seconds correct to one-tenth of a second, the number of responses given, and the latencies. Each subject's test sheets told which items he had answered and which ones were correct. Moore (1969) gives a complete description of experimental equipment and the organization of trials. The subjects completed 194 trials and recorded 5,952 latencies.

Results

The Poisson process applies with varying degrees of success to three of the six highly speeded tests in the battery and the Erlang process applies to one test. For two tests (Maze Tracing and Addition) the stochastic models had no application. We first discuss the results of the most successful of the Poisson model tests.

Finding A's is a clerical speed test of two minutes' duration. In each column of 41 words the task is to check the five words having the letter "a." The score is the number of words correctly checked. Exponential distributions fit the latencies of individual subjects and each subject's number of responses is a realization of a Poisson random variable. Finding A's satisfies the necessary and sufficient conditions for a Poisson process.

Each subject completed four trials in Finding A's. Six of the ten subjects have estimates of one for the test parameter, n, in three or four of their trials and over the four trials combined estimates of n equalled one for nine subjects. For each of the four parts of the test, estimates of n

equalled one and for the performance of all ten subjects on the whole test the estimate of n was again one.

Table 2 shows observed frequencies of latencies in transformed time (equation 21) and the theoretical frequencies computed from the exponential column of Table 1. The class intervals of Table 2 are the same as those of Table 1. When observed frequencies for the ten subjects are combined into a single distribution and expected frequencies are likewise combined, the χ^2 goodness-of-fit accepts the null hypothesis at the 10% level.

The test of Poisson distributiveness of each subject's observed score requires a well-known theorem in Poisson variables: If X_n are independent Poisson random variables with parameters Ω_n, then ΣX_n has a Poisson distribution with parameter $\Sigma\Omega_n$ (Lord and Novick, 1968, p. 483); and, conversely in the stochastic context of a Poisson process, ΣX_n partitions into n independent Poisson random variables with parameters proportional to the length of the time intervals. Let $(o,t_1^*],(t_1^*,t_2^*],\ldots,(t_{n-1}^*,t_n^*]$ be n equal partitions of the total transformed time for subject i and assume the Poisson mean value functions of random variables in these intervals to be

$$(40) \qquad \omega = \Omega(o,t_1^*) = \Omega(t_1^*,t_2^*) = \ldots = \Omega(t_{n-1}^*,t_n^*) \quad .$$

The relationship between equations (3) and (40) is

$$(41) \qquad \Omega_{iv} = n\omega \quad .$$

Table 2

Observed and Expected Frequencies for Latencies in Transformed Time Units for Ten Subjects in Finding A's. Data Grouped as in Table 1

1		2		3		4		5		6		7		8		9		10	
f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e
28	23	34	33	30	38	38	42	34	35	14	17	42	35	19	31	19	21	20	22
21	18	27	26	23	30	33	33	26	27	15	13	23	27	26	24	20	17	11	17
13	16	21	20	33	23	19	26	14	21	11	11	23	21	24	19	13	13	18	14
5	10	17	16	21	18	23	20	10	16	9	8	13	16	19	14	5	10	9	10
9	9	7	12	12	14	17	15	17	13	5	6	16	13	9	11	9	8	12	8
7	7	13	10	13	11	19	12	18	10	5	5	8	10	8	8	10	6	7	6
4	5	7	7	11	9	4	9	4	7	4	4	3	7	8	7	3	5	5	5
3	4	6	6	6	7	15	8	7	6	1	3	5	6	5	5	3	4	3	4
7	6	5	8	13	10	4	10	16	8	6	4	11	8	10	7	4	5	11	5
3	3	6	5	8	6	4	6	5	5	0	3	9	5	7	4	5	3	1	3
4	2	3	3	2	4	5	4	3	3	2	2	1	3	1	3	1	2	0	2
1	1	2	2	1	2	3	2	1	2	0	1	1	2	1	2	1	1	3	1
0	1	1	2	0	2	1	2	1	2	4	1	1	2	0	2	2	1	0	1
0	1	0	1	1	1	2	1	0	1	1	0	0	1	0	1	1	0	0	1
0	0	1	0	0	0	2	0	0	0	1	0	0	0	1	0	1	0	0	0
0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
105	106	150	151	174	175	191	190	156	156	78	78	156	156	138	138	97	96	100	100

Note: Each subject's four trials are combined into one continuous trial.

Let the number of responses given by subject i in the interval $(t^*_{j-1}, t^*_j]$ be X_{ij} for $j = 1,2,\ldots,n$ and assume the X_{ij} to be Poisson distributed with common mean value function ω. The test of this assumption uses the Poisson index of dispersion,

$$(42) \qquad I_p = \sum_{j=1}^{n} \frac{(X_{ij} - \bar{x}_i)^2}{\bar{x}_i}$$

which is distributed as χ^2 (Snedecor and Cochran, 1967, p. 232).

Table 3 shows the number of responses given by each subject in consecutive intervals of 10 units of transformed time. The χ^2 tests (Table 4) reveal that we may accept the observed scores of nine subjects as being realizations of Poisson random variables.

The Poisson model predicts for each subject that each unit of transformed time should contain one response. Forty per cent of the entries in Table 3 are 9, 10 or 11 and from Table 4 we find that 6 of the 10 subjects have means in the range 9.5 to 10.5. The grand mean for the ten subjects is 9.7. The model also predicts that the length of the test in transformed time is an estimate of an examinee's true score, Ω_{iv}. From the properties of the Poisson function, Y_{iv} is an estimate of Ω_{iv} (equation 4). Table 5 lists these two estimates for each subject. The t-test for differences between matched pairs enables us to retain the null hypothesis of no difference between the estimates.

Two other speeded tests to which the Poisson process has some application are Hidden Patterns and Division. Hidden Patterns is a perceptual speed test

Table 3

Number of Responses Recorded by Ten Subjects in Each Ten Unit Interval of Transformed Time of Finding A's

Class Intervals (transformed time)	Subjects (Number of responses) 1	2	3	4	5	6	7	8	9	10	Total
0-10	13	14	11	12	16	11	13	11	13	12	126
10-20	10	9	10	13	8	13	11	5	8	9	96
20-30	12	9	5	4	9	7	10	9	10	12	87
30-40	8	13	12	8	7	8	9	9	11	9	94
40-50	12	10	12	7	11	9	15	9	11	14	110
50-60	10	10	8	10	12	7	9	13	12	8	99
60-70	9	14	12	14	16	8	18	11	8	12	122
70-80	16	9	11	1	8	7	10	15	7	10	94
80-90	8	11	5	14	11	4	5	6	4	7	75
90-100	7	11	15	5	10		11	9	8	7	83
100-110		11	13	4	10		11	12			61
110-120		7	10	11	7		12	11			58
120-130		11	11	7	8		13	9			59
130-140		8	9	15	8		9	6			55
140-150			6	8	5						19
150-160			7	9	7						23
160-170			9	10							19
170-180			8	10							18
180-190				5							5
190-200				10							10
200-210				7							7
210-220				5							5

Note: Each subject's four trials are combined into one continuous trial.

Table 4

Calculations Based on Data from Table 3

Description	Symbols	Subjects 1	2	3	4	5	6	7	8	9	10
Number of responses	$\sum X$	105	147	174	189	153	74	156	135	92	100
Number of class intervals	N	10	14	18	22	16	9	14	14	10	10
Mean number of responses	$\overline{X} = \frac{\Sigma X}{N}$	10.5	10.5	9.7	8.6	9.6	8.2	11.1	9.6	9.2	10.0
Squared deviations	$\sum (X-\overline{X})^2$	68.5	57.5	132.0	291.3	143.9	53.6	123.7	101.2	65.6	52.0
Poisson index of dispersion	$\sum \frac{(X-\overline{X})^2}{\overline{X}}$	6.524	5.476	13.655	33.910	15.053	6.514	11.103	10.496	7.130	5.200
Probability level of chi-square*	χ^2	.7	.95	.7	.05	.4	.6	.6	.6	.6	.8

*Based on (N - 1) degrees of freedom. Probability levels are approximated to one decimal place in most cases.

Table 5

Significance Test for the Differences between Two Independent Estimates of the Mean Value Function for the Poisson Distribution of Responses on Finding A's

Subject Number	First estimate of Ω_{iv} (Sum of latencies in transformed time)	Second estimate of Ω_{iv} (Number of responses)	Difference D
1	98.5	105	-6.5
2	145.0	150	-5.0
3	178.3	174	+4.3
4	225.3	191	+34.3
5	163.6	156	+7.6
6	92.8	78	+14.8
7	141.3	156	-14.7
8	142.8	138	+4.8
9	105.3	97	+8.3
10	99.7	100	-0.3
Totals	1392.6	1345	$\sum D$ = +47.6
			$\bar{D}$ = +4.76
			$S_{\bar{D}}$ = 4.25
			t = 1.120
			d.f. = 9
			$t_{.05}$ = 2.262

of two minutes' duration. Each item consists of a geometrical pattern and in some items a given configuration, λ, is embedded. The task is to select the items having the given configuration. Six subjects have latency distributions showing only minor departures from exponential distributions. However, the combined latency distribution of all subjects differs significantly on a χ^2 test from the theoretical distribution. The Poisson index of dispersion shows that the observed scores of eight subjects are realizations of Poisson random variables.

The partitions of transformed time into 10 unit intervals do not yield numbers of responses in Hidden Patterns clustering about 10 to the same extent as they do for Finding A's. Only 26% of partitions have 9, 10 or 11 responses and the means of cell entries range from 8.6 to 9.8. The t-test for differences between the length of the test in transformed time and the number of responses for the ten subjects is significant at the 0.1% level. That is, the two estimates of true score differ significantly. The Poisson model does not fit Hidden Patterns with the same precision it does for Finding A's. The notion of a latency being the sum of a random variable and a constant may be inadequate (equation 8). To retain the simplicity of a linear model we could suppose that the latency for item j is the sum of two random variables and a constant,

$$t_j = c_i + \tau_j + \tau'_j \quad . \tag{43}$$

If an additional random variable, τ'_j, exists then our model would introduce errors into the estimates of λ_i and c_i which would bias the transformation of time. Moore (1969) has done some work on latencies structured as equation (43), but the problem of biased estimates of λ_i and c_i remains.

Division is a speeded computation test of dividing 2- or 3-digit numbers by single digit numbers. For five subjects distribution of latencies fit theoretical exponential distributions very well. The χ^2 tests on Poisson indices of dispersion accept the observed score of each subject as the realization of a Poisson random variable. The t-test for differences between two independent estimates of Ω_{iv} is not significant, but the standard error of the difference was relatively large--it exceeded four times the magnitude of the standard error for Finding A's (Table 5).

The Poisson model fits Division at a level of precision less than it did for Finding A's but greater than for Hidden Patterns. However, to be an adequate theory for the Division test, the model needs modification. As with Hidden Patterns, the assumed structure of latencies, equation (8), does not accommodate all the systematic variation in latencies.

The speeded test of Subtraction and Multiplication met the requirements of the Erlang process as a model for the performance of individual examinees. This test alternates 10 items of subtracting 2-digit numbers from 2-digit numbers and 10 items of multiplying 2-digit numbers by single-digit numbers. The two parallel forms each have 60 items and subjects completed 3-minute trials on each form. For 11 of the 20 trials the estimates of the test parameter were $n = 2$. For 3 trials in which $n = 3$, exact estimates slightly exceeded 2.5; and for one trial in which $n = 1$, the exact estimate is slightly less than 1.5. Forms 1 and 2 each have overall estimates of $n = 2$. Seven of the ten subjects have $n = 2$ for the combination of both forms of the test.

By substituting $n = 2$ into equation (39), the theoretical density for latencies in this test has the gamma form,

$$(44) \qquad \phi(L_{ij}|i) = \lambda_i(L_{ij} - c_i)e^{-\lambda_i(L_{ij}-c_i)} \qquad L_{ij} \geq c_i$$

$$= 0 \qquad L_{ij} < c_i \ .$$

The third column of Table 1 displays probabilities calculated from equation (44). Table 6 shows observed frequencies of latencies in transformed time and theoretical frequencies computed from Table 1. Class intervals are the same as for Table 1. When observed frequencies are combined into a single distribution and expected frequencies likewise combined, the χ^2 goodness-of-fit test accepts the null hypothesis at the 40% level.

To test the Erlang distributiveness of each subject's observed score requires a slightly more complicated version of the strategy developed for the Poisson process. The theoretical Erlang distribution is equation (36). The two terms on the right of equation (36) are each a form of the Poisson function. Rewritten entirely as Poisson probabilities equation (36) becomes

$$(45) \qquad \Pr(Y_{iv} = y|\Omega^*_{iv}) = \Pr(Y_{iv} = 2y|\Omega^*_{iv}) + \Pr(Y_{iv} = 2y + 1|\Omega^*_{iv}) \quad .$$

As before, we partition the total transformed time for a trial into n equal intervals $(o, t^*_1], (t^*_1, t^*_2], \ldots, (t^*_{n-1}, t^*_n]$. Let the random variables representing numbers of responses in these intervals be $Y_{i1}, Y_{i2}, \ldots, Y_{in}$. The random variables $2Y_{i1}, 2Y_{i2}, \ldots, 2Y_{in}$ are mutually independent and Poisson distributed with mean value function Ω^*_{iv}/n in each case. The random variables $(2Y_{i1} + 1), (2Y_{i2} + 1), \ldots, (2Y_{in} + 1)$ are mutually independent

Table 6

Observed and Expected Frequencies in Transformed Time Units for Ten Subjects in Subtraction and Multiplication

1		2		3		4		5		6		7		8		9		10	
f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e	f_0	f_e
2	3	0	2	1	3	1	3	6	3	1	2	2	3	3	3	2	2	5	3
4	6	2	4	6	7	7	7	7	7	2	4	5	7	7	7	4	5	1	7
12	9	10	6	12	10	6	10	8	9	7	6	8	9	6	10	5	6	8	10
8	9	5	6	9	11	16	11	6	10	9	6	12	10	10	11	10	7	14	11
7	9	5	6	3	10	9	10	13	10	6	6	12	9	18	10	6	7	11	10
10	9	6	6	12	10	9	11	11	10	3	6	11	10	7	11	7	7	10	10
9	8	2	5	6	9	12	10	9	9	8	6	7	9	9	10	4	6	11	10
5	7	4	5	13	8	8	8	7	8	3	5	5	7	10	8	5	5	12	8
15	13	11	8	11	14	16	15	10	13	7	8	10	13	16	15	10	9	10	14
6	9	7	6	19	10	12	10	12	10	9	6	16	9	6	10	8	7	10	10
7	7	7	4	2	7	7	8	13	7	7	4	9	7	6	8	6	5	4	7
6	5	2	3	4	5	11	5	3	5	3	3	6	5	6	5	3	3	4	5
6	5	3	3	10	6	3	6	2	6	3	4	5	6	6	6	5	4	11	6
2	2	2	2	4	3	3	3	5	3	0	2	0	2	6	3	0	2	4	3
3	1	1	1	1	2	0	2	0	2	1	1	0	2	2	2	0	1	1	2
0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
102	102	67	67	114	115	120	119	112	112	69	69	108	108	119	119	75	76	116	116

and Poisson distributed with mean value function Ω^*_{iv}/n in each case. Letting X_{ij} be the realization of random variables Y_{ij}, the Poisson index of dispersion for each subject is

$$(46) \qquad I_p = \frac{16}{4\bar{x} + 1} \sum_{j=1}^{n} (X_{ij} - \bar{x})^2 \quad .$$

The entries in the body of Table 7 are X_{ij} with transformed time partitioned into intervals of 10 units. The χ^2 tests in Table 8 indicate that 8 subjects have observed scores acceptable as realizations of Erlang random variables.

The Erlang model predicts for each subject that each ten units of transformed time should contain 4.75 items (equation 38). Sixty three per cent of entries in Table 7 are 4, 5 or 6. Eight of the ten subjects have means within the range 4.75 ± 0.5 and the remaining two subjects have means of 4.1 and 5.3. The unweighted mean for the ten subjects is 4.86.

The model also predicts that Ω^*_{iv} from the Erlang distribution equals the length of the test in transformed time. From observed scores, the estimate of Ω^*_{iv} is

$$(47) \qquad \Omega^*_{iv} = (4 \sum_{j=1}^{n} X_{ij} + n)/2$$

where n is, as before, the number of partitions of transformed time. Table 9 lists the two independent estimates of Ω^*_{iv} for each subject and the resulting t-test retains the null hypothesis of no difference between the pairs of estimates. The difference column in Table 9 shows that the

Table 7

Number of Responses Recorded by Ten Subjects in Each Ten Unit Interval of Transformed Time of Subtraction and Multiplication

Class Intervals (transformed time)	Subjects 1	2	3	4	5	6	7	8	9	10	Total
	(number of responses)										
0-10	7	5	7	7	8	7	9	8	7	3	68
10-20	5	4	6	6	6	5	3	4	5	7	51
20-30	5	7	4	6	5	6	4	3	5	4	49
30-40	3	4	4	3	2	4	3	5	4	6	38
40-50	5	4	5	5	4	6	7	2	5	3	46
50-60	4	4	4	4	3	2	5	6	4	3	39
60-70	5	5	5	4	3	6	4	5	6	4	47
70-80	7	3	5	3	7	5	5	7	7	5	54
80-90	4	4	1	9	6	6	10	4	4	6	54
90-100	4	8	4	3	4	3	3	5	5	6	45
100-110	7	4	5	4	7	4	9	3	4	5	52
110-120	3	4	5	5	6	7	5	6	7	7	55
120-130	6	3	6	7	4	5	5	5	5	7	53
130-140	4	8	4	5	6		3	5	4	2	41
140-150	5		4	6	5		7	3		8	38
150-160	4		5	5	5		5	6		4	34
160-170	3		4	3	5		3	6		3	27
170-180	6		5	5	5		4	8		6	39
180-190	5		3	5	7		6	2		6	34
190-200	6		4	5	3		6	6		6	36
200-210	3		4	9	8			7		4	35
210-220			5	2	4			6		7	24
220-230			6	5				2		3	15
230-240			3	5				1		5	14
240-250			5					4			9

Note: Each subject's two trials are combined into one continuous trial.

ERIC Full Text Provided by ERIC

Table 8

Calculations Based on Table 7

Description	Symbols	Subjects 1	2	3	4	5	6	7	8	9	10
Number of responses	$\sum X$	101	67	112	120	113	66	106	119	72	120
Number of class intervals	N	21	14	25	24	22	13	20	25	14	24
Mean number of responses	$\overline{X} = \frac{\Sigma X}{N}$	4.8	4.1	4.5	5.0	5.1	5.1	5.3	4.6	5.1	5.0
Poisson index of dispersion	eq. 46	28.896	29.574	38.311	56.000	44.541	20.827	64.468	71.954	13.417	49.905
Probability level of chi-square*	X^2	.9	.3	.9	.3	.4	.7	.01	.025	.98	.4

*Based on (2N - 1) degrees of freedom. Probability levels are approximated to the nearest tenth.

Table 9

Significance Test for the Differences between Two Independent Estimates of the Parameter of the Erlang Distribution of Responses for Subtraction and Multiplication

Subject Number	First estimate of Ω^*_{iv} (Sum of latencies in transformed time)	Second estimate of Ω^*_{iv} (Summed linear function of responses)	Difference D
1	216.9	212.5	+4.4
2	139.3	141.0	-1.7
3	251.5	236.5	+15.0
4	240.0	252.0	-12.0
5	219.1	235.0	-25.9
6	137.2	138.5	-1.3
7	207.6	220.0	-12.4
8	252.5	248.5	+4.0
9	143.5	151.0	-7.5
10	239.6	244.0	-4.4
Totals	2047.2	2079.0	$\sum D=$ -41.8
			$\bar{D}=$ -4.18
			$S_{\bar{D}}=$ 3.55
			$t=$ 1.177
			$t_{.05}=$ 2.262

greatest difference is approximately 10% of the magnitude of the two estimates, and that all remaining differences are 5% or less of the magnitude of their two respective estimates. By theory of the Erlang model, each subject's true score is $0.475\Omega^*_{iv}$.

The case study of the performances of ten subjects in Subtraction and Multiplication shows that latencies are gamma distributed; that for 8 subjects the number of responses recorded in fixed time are realizations of Erlang random variables; and that the convolution of latencies over the full test time has as its argument the parameter of the Erlang distribution.

The failure of two subjects to survive the χ^2 tests in Table 8 is the only blemish preventing complete acceptance of the Erlang model as a validated theory for these ten subjects on Subtraction and Multiplication.

Conclusions

The stochastic models have met with varying degrees of success for the six highly speeded tests used in this study. Maze Tracing and Addition reject the models completely. Hidden Patterns and Division accept the Poisson model partially. Finding A's accepts the Poisson model fully, and Subtraction and Multiplication accepts an Erlang model with only minor reservation.

The relationships of the models to the three speeded arithmetic tests are interesting. The research does not clarify whether the relationship results from the particular tests, or from mental processes, but we suspect the latter. The design of the tests is careful and expert, and they are valid instruments for measuring the skills and abilities they intend to measure.

The partial failure of the models for Hidden Patterns and Division is probably the result of an inadequate transformation function for time. Many options exist for transformation functions for time. Further research may well produce a function which will enable the models to act successfully for Hidden Patterns and Division.

References

Bush, R. R. and Mosteller, F. Stochastic models for learning. New York: Wiley, 1955.

French, J. W., Ekstrom, R. B., and Price, L. A. Kit of reference tests for cognitive factors. Princeton, N. J.: Educational Testing Service, 1963.

Haight, F. A. Handbook of the Poisson distribution. New York: Wiley, 1967.

Karlin, S. A first course in stochastic processes. New York: Academic Press, 1966.

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

Meredith, W. The Poisson distribution and the Poisson process in psychometric theory. Research Bulletin 68-42. Princeton, N. J.: Educational Testing Service, 1968.

Moore, W. E. Stochastic true-score models for highly speeded mental tests. Unpublished doctoral dissertation, Harvard University, 1969.

Novick, M. R. The axioms and principal results of classical test theory. Journal of Mathematical Psychology, 1966, 3 (1), 1-18.

Parzen, E. Stochastic processes. San Francisco: Holden-Day, 1962.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nelson and Lydicke, 1960.

Snedecor, G. W. and Cochran, W. G. Statistical methods. (6th ed.) Ames: Iowa State University Press, 1967.