AFOSR 68-2155

AD 677203

1. This document has been approved for public release and sale; its distribution is unlimited.

THE SHUFORD-MASSENGILL CORPORATION

SM

57

R-5

THE EFFECT OF GUESSING ON THE QUALITY OF PERSONNEL AND COUNSELING DECISIONS

EMIR H. SHUFORD, Jr. and H. EDWARD MASSENGILL

THE FIRST SEMIANNUAL TECHNICAL REPORT (WHICH COVERS THE PERIOD MAY 1966 THROUGH OCTOBER 1966) OF WORK PERFORMED UNDER CONTRACT NUMBER AF 49 (638) - 1744, ARPA ORDER NUMBER 833, BY THE SHUFORD - MASSENGILL CORPORATION, P.O. BOX 26, LEXINGTON, MASSACHUSETTS, 02173

DECISION-THEORETIC PSYCHOMETRICS: AN INTERIM REPORT, NOVEMBER 1966

Emir H. Shuford, Jr. and H. Edward Massengill

ABSTRACT

In Section A, A Logical Analysis of Guessing, appropriate test-taking strategies are derived for six major test-scoring procedures. Three commonly used definitions of guessing are interpreted as corresponding degree-of-confidence distributions. The ability of the testing procedures to separate these distributions from those representing higher degrees of knowledge is considered with the major result that only admissible probability measurement performs satisfactorily.

In Section B, The Effect of Guessing on the Quality of Personnel and Counseling Decisions, the fundamental probability distributions for total test scores are derived by assuming that each person knows the answers to some items and guesses on the remaining items. Analysis of a 10-item test shows that guessing levels encountered in practice (a) seriously degrade the value of selection, placement, and counseling decisions, (b) significantly impair test reliability and validity, and (c) magnify the influence of testwiseness.

In Section C, The Worth of Individualizing Instruction, equations are developed for expressing the cost and gain for applying an instructional sequence. The expected return from assigning instruction on the basis of (1) admissible probability measurement, (2) admissible choice testing, (3) conventional choice testing, (4) prior information only, and (5) matching the average student is computed for each of seven distributions of state of knowledge. The performance of (1) is outstanding; that of (2), (3), and (4) is disappointing, while (5) does surprisingly well.

B. The Effect of Guessing on the Quality of Personnel and Counseling Decisions

A major potential domain of application of admissible procedures is to obtain test information to guide personnel decisions such as selection and classification. Another major domain is to obtain test data to inform counseling decisions and recommendations. It should be recognized that the decisions involved in these two domains have some important differences, for example, selection and classification decisions are clearly institutional decisions and the utility to the institution can often be approximated by a linear function of the true ability of a selected individual. In counseling, on the other hand, the emphasis is on giving advice and recommendations to the individual and part of this advice concerns an estimate of his true ability level. Here, good advice is accurate advice and the consequences of an error can be taken as proportional to the square of the difference between the individual's estimated ability level and his true ability level. While the decision problems in these two domains of application typically involve different utility functions, they have a very important similarity. These decisions are based upon an individual's total score which is taken to be an indicant of his ability level. Therefore, an analysis of factors affecting total test score can serve as the basis for estimating the effectiveness of both personnel and counseling decisions.

The effect of guessing and of test-wiseness are two interesting problems in the theory and practice of testing. Though these problems have never been resolved, they are typically ignored especially in practice. Now, it should be intuitively evident from Section A above that conventional choice testing and its modifications cannot detect guessing. Furthermore, under the conditions of testing specified below, it is mathematically proven that no analysis of information internal to the conventional choice test can detect the extent of guessing. These observations lead to the conjecture that the problem of guessing is ignored in practice because conventional choice testing is incapable both of preventing guessing and of detecting the presence of guessing. However, there is another conjecture possible, certainly one that should be considered. This conjecture is that guessing has no singificant effect on personnel or counseling decisions and, thus, can and should be ignored in practice. Now, the coming into being of the new admissible procedures certainly makes it possible to decide these issues. First, and most importantly, admissible procedures make it possible to eliminate the effect of guessing in an objective or semi-objective examination as should be clear from Section A above. Thus, empirical comparisons of the performance of admissible tests with that of conventional choice tests should be able to settle the issue. Additionally, however, the mere fact that guessing has now been clearly defined and can be empirically measured makes it possible to use new operational definitions in a formal explication of test-taking behavior. This allows us to mathematically analyze different testing situations and to predict the effect of guessing in a wide variety of situations. This analysis and the resulting prediction should be quite useful in guiding decisions concerning the substitution of admissible procedures which eliminate guessing for conventional testing procedures. The remainder of this section begins such an analysis.

To return to the issue of test-wiseness and to anticipate some of the later results, whether or not an individual chooses to guess at those items which he does not know can make a considerable difference in his test score. Thus, we can expect that individuals with a great deal of experience taking conventional choice tests will learn to guess and, if possible, never skip an item. This individual test-taking strategy, of guessing at all the items which one does not know rather than refusing to guess at these items and just skipping them, we identify as test-wiseness. The mathematical formulation makes it possible to compare the performance of test-wise individuals with those who are not test-wise and to predict the effect of this individual difference under a wide variety of conditions.

THE FORMAL MODEL

Assume that there exists, at least conceptually, a rather large pool of test items and a population of persons who will eventually take the test. In principal, it is possible to conceive of having all the persons take all of the test items, and that instead of being given a conventional choice or constructed-response test, the persons take this super-test by using an admissible probability measurement or an admissible choice procedure. Data obtained through the use of these procedures would indicate whether a person was (a) well-informed, (b) relatively uncertain and possibly guessing, or (c) misinformed with respect to the answer to each test item. Assume for the sake of simplicity, and not too unrealistically in the case of certain types of tests, that the persons either pretty well know the answer to the test item or they are uncertain as to the answer, so we can now characterize each person as knowing a proportion, p, of the test items and being uncertain about the rest of the items. Suppose further that this uncertainty were such that if any person were given a conventional choice test, he would guess the correct answer to the item with a certain constant probability, θ, of being correct. This is the essence of the basic model. There is a population of test items; there is a population of persons. Each person knows the answer to a certain proportion of the test items. This proportion corresponds to ability, achievement, or true score, in the sense that it is the one-dimensional quantity which determines the effectiveness of decisions based upon testing information. The remainder of the items the person guesses with constant probability, θ, of getting a correct answer.

Thus far, discussion has been in terms of a super-test based on all items in the pool. Any test actually administered can be viewed as a random selection of the samples of items from this pool. Let n represent the number of items in this actual test. Now take θ to be zero. No person guesses at any of the items. A person's score, x, on this test is equal to the number of items that he answered correctly. It depends both upon the number of items, n, in the test and upon the proportion, p, of items in the population of test items that the person knows. Since the items in the actual test have been randomly sampled from the items in the pool, the person's test score, x, is a random variable with a binomial distribution and can be written as

$$(1) \qquad f_b(x|p,n) = \binom{n}{x} p^x (1-p)^{n-x}.$$

This is the distribution of the pupil's score given that p is known. However, if p were known, there would be no point in giving the test since the purpose of obtaining the test score is to obtain information about p. The decision maker and user of the test information must have some information about p prior to observing the test score for a person. Prior information about p can most conveniently be represented by a Beta distribution over the interval, [0,1].

$$(2) \qquad f_\beta(p|a,b) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad a,b > 0.$$

Now, in the case of no guessing, $\theta = 0$, choice testing can be represented by the well-known Bernoulli process and the many results of applied statistical decision theory (Raiffa and Schlaifer, 1961) can be applied with ease. For example, the marginal or unconditional distribution of the test score, x, is a Beta-binomial distribution,

$$(3) \qquad f_{\beta b}(x|a,b,n) = \int_0^1 f_b(x|p,n) f_\beta(p|a,b)\,dp$$
$$= \frac{(x+a-1)!\ (n-x+b-1)!\ n!\ (a+b-1)!}{x!\ (a-1)!\ (n-x)!\ (b-1)!\ (n+a+b-1)!}$$

The posterior, or conditional distribution of p, given x, is, like the prior, a

Beta distribution, but with parameters, x+a and n-x+b,

$$f_\beta(p|x+a,n-x+b) = \frac{1}{B(x+a,n-x+b)} p^{x+a-1}(1-p)^{n-x+b-1} \tag{4}$$

and with mean, $(x+a)/(n+a+b)$, and variance, $(x+a)(n-x+b)/(n+a+b)^2(n+a+b+1)$.

Characteristics of this Bernoulli process have been extensively analyzed for many decision problems and the results are relatively tractable. Thus, if there were no guessing occurring in testing, there would be available an extensive literature containing many results which could be immediately translated into the terminology of test theory and used as the basis for a decision-theoretic psychometrics dealing with institutional decisions. Guessing does occur, however, in conventional testing and we must take this into account. In doing so, the mathematics becomes much less tractable and we must leave behind most of the neat, analytic equations of the Bernoulli process. Allowing for the possibility of guessing during the test-taking process yields equations which are not readily integrated. Therefore, there is no sacrifice in getting rid of the one continuous distribution by using a discrete density function to approximate the distribution in (2) expressing the distribution of ability levels in the population of persons. Though, in later work we will consider different distributions of ability level, in this report we use the distribution shown graphically in Figure 1 and given numerically at the bottom of Table 2. It is a symmetric distribution with mean equal to one-half and represents tests of average difficulty.

Now, let us analyze a ten-item test. Later work will consider both shorter and longer tests, but a ten-item test is sufficiently long to bring out the effects of guessing and test-wiseness, but not so long as to make the presentation of the computational techniques unbearable. The initial distribution, (See Figure 1) allows for nine different ability levels with p ranging from .1 to .9 in steps of .1. Thus, with no guessing, the conditional distributions of test scores are binomial according to (1) and are given in Table 1.

According to the definition of conditional probability, $P(AB) = P(A|B)P(B)$, the joint probabilities of x and p are obtained by multiplying each conditional probability of x by the appropriate marginal probability of p and are shown in Table 2. Summing over the rows of Table 2 yields the marginal distribution of x also shown in Table 2. The joint probabilities given in this table contain all the information about the testing process itself.

Now suppose that a person guesses at the answer to each item that he doesn't

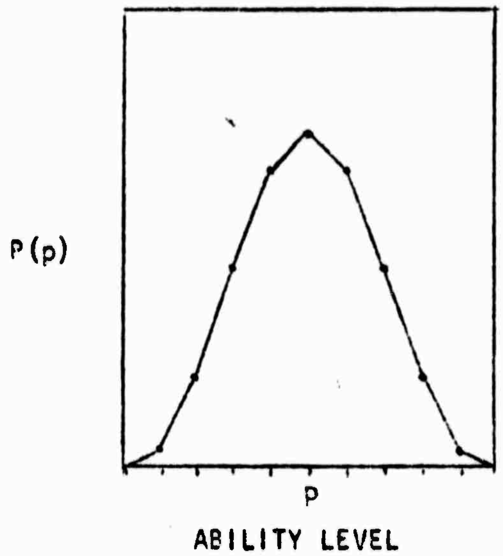


Figure 1. Marginal distribution of ability level.

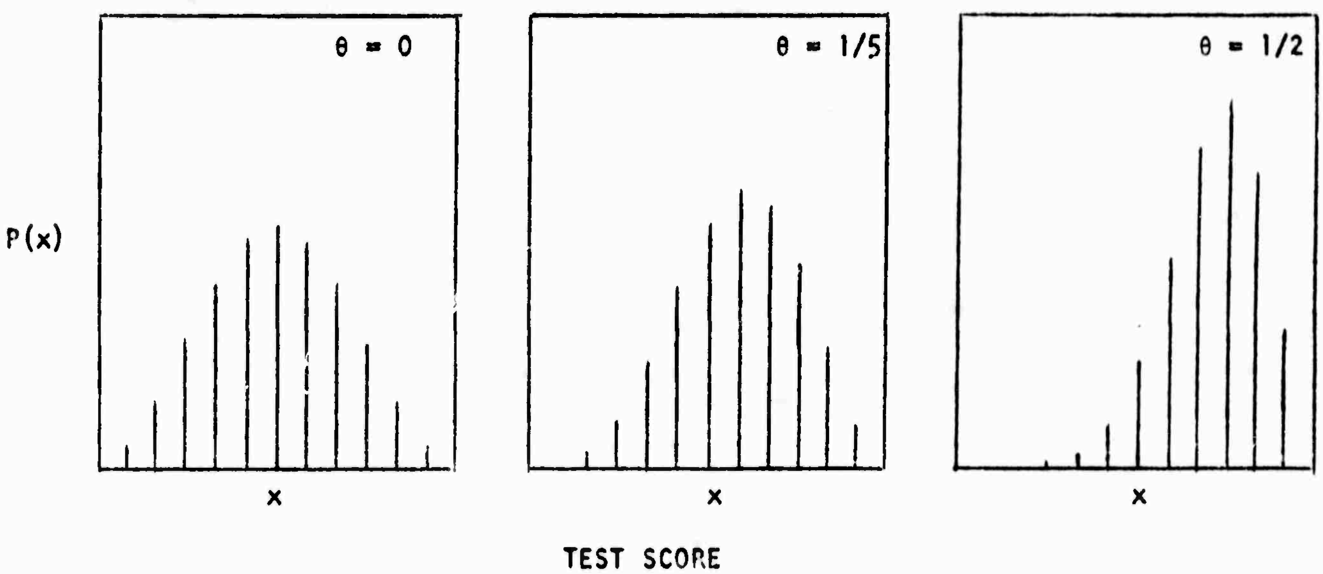


Figure 2. Marginal distributions of test score for a 10-item test affected by different degrees of guessing.

Table 1.

Conditional distributions of x given p. No guessing ($\theta = 0$).
Entries to be scaled times 10^{-5}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10			1	10	98	605	2825	10737	34868
9			13	158	976	4031	12106	26844	38742
8		8	145	1061	4395	12093	23347	30199	19371
7	1	78	900	4247	11719	21499	26683	20133	5739
6	14	551	3676	11148	20507	25082	20012	8808	1117
5	148	2642	10292	20066	24610	20066	10292	2642	148
4	1117	8808	20012	25082	20507	11148	3676	551	14
3	5739	20133	26683	21499	11719	4247	900	78	1
2	19371	30199	23347	12093	4395	1061	145	8	
1	38742	26844	12106	4031	976	158	13		
0	34868	10737	2825	605	98	10	1		

Table 2.

Joint and marginal distributions of x and p. No guessing ($\theta = 0$).
Entries to be scaled times 10^{-5}.

Score (x)	.1	.2	.3	.4	.5	.6	.7	.8	.9	Total P(x)
10				2	21	117	366	616	356	1478
9			2	31	214	780	1569	1539	395	4530
8			19	205	961	2340	3026	1731	198	8480
7		4	117	822	2563	4160	3459	1154	59	12338
6		32	477	2157	4485	4853	2594	505	11	15114
5	1	152	1334	3882	5382	3882	1334	152	1	16120
4	11	505	2594	4853	4485	2157	477	32		15114
3	59	1154	3459	4160	2563	822	117	4		12338
2	198	1731	3026	2340	961	205	19			8480
1	395	1539	1569	780	214	31	2			4530
0	356	616	366	117	21	2				1478
Total P(p)	1020	5733	12963	19349	21870	19349	12963	5733	1020	100000

know and that his probability of getting the correct answer is θ, for each of these items. The person knows the answer to r of the items; he guesses the answer to each of the remaining n-r items. Given the number of items, r, that the person knows, the distribution of the number of items, t, that the person guesses correctly is binomial with parameters, θ and n-r,

$$P(t|r,n,\theta) = f_b(t|\theta,n-r) = \binom{n-r}{t}\theta^t(1-\theta)^{n-r-t}. \tag{5}$$

These distributions are shown in Table 3 for $\theta = 1/5$ and in Table 4 for $\theta = 1/2$. Note that these are probably the extreme values that θ can assume in conventional choice testing. With $\theta = 1/5$ representing the lowest possible guessing probability for a five-alternative multiple choice test and with $\theta = 1/2$ representing the largest possible guessing probability which may be encountered in any multiple-choice or constructed-response test.

These tables (Tables 3 and 4) are arranged as they are in order to make it clear that guessing adds to the score due to the person's ability level and that a particular test score, x, may arise in a number of ways corresponding to different combinations of r and t which sum to x. For example, a person may obtain a test score of 2, by knowing none of the items but successfully guessing two of them, knowing one of the items and successfully guessing one of them, or by knowing two items. These guessing distributions are conditional upon r. The distribution of x conditional upon p may be found by multiplying the conditional probability of r times the probability of t and summing over those values that yield the same x as shown in Equation 6 below.

$$f(x|p,n,\theta) = \sum_{r=0}^{x}\binom{n}{r}p^r(1-p)^{n-r}\binom{n-r}{x-r}\theta^{n-r}(1-\theta)^{n-x} \tag{6}$$

This equation could be used to obtain the conditional distributions of x given p. These rather extensive computations may be avoided, however, by making use of the theorem given below:

THEOREM 1. If $x = r + t$ where the distribution of r is binomial with parameters, p and n, and the distribution of t is binomial with parameters, θ and n-r,

$$f(x|p,n,\theta) = f_b(x|p+\theta(1-p),n).$$

Table 3.

Guessing distributions conditional upon the number of test items known. Guessing probability equal to 1/5. Entries to be scaled times 10^{-5}.

No. of Items Known (r)	Test Score (x) 0	1	2	3	4	5	6	7	8	9	10
10											100000
9										80000	20000
8									64000	32000	4000
7								51200	38400	9600	800
6							40960	40960	15360	2560	160
5						32768	40960	20480	5120	640	32
4					26214	39322	24576	8192	1536	154	6
3				20972	36700	27525	11469	2867	430	36	1
2			16777	33555	29360	14680	4587	918	115	8	
1		13422	30199	30199	17616	6606	1651	276	29	2	
0	10737	26844	30199	20133	8808	2642	551	78	8		

Table 4.

Guessing distributions conditional upon the number of test items known. Guessing probability equal to 1/2. Entries to be scaled times 10^{-5}.

No. of Items Known (r)	Test Score (x) 0	1	2	3	4	5	6	7	8	9	10
10											100000
9										50000	50000
8									25000	50000	25000
7								12500	37500	37500	12500
6							6250	25000	37500	25000	6250
5						3125	15625	31250	31250	15625	3125
4					1563	9375	23437	31250	23437	9375	1563
3				781	5469	16406	27344	27344	16406	5469	781
2			391	3125	10937	21875	27344	21875	10937	3125	391
1		195	1758	7031	16407	24609	24609	16407	7031	1758	195
0	98	976	4395	11719	20507	24610	20507	11719	4395	976	98

Proof: By rewriting the binomial coefficients and rearranging terms in Equation 6 we obtain

(7a) $$f(x|p,n,\theta) = \sum_{r=0}^{x} \frac{n!}{r!(n-r)!} \frac{(n-r)!}{(x-r)!(n-x)!} p^r(1-p)^{n-r}\theta^{x-r}(1-\theta)^{n-x} .$$

Now, isolate the terms not dependent upon the variable of summation, r.

(7b) $$f(x|p,n,\theta) = \frac{n!}{(n-x)!} (1-\theta)^{n-x} \sum_{r=0}^{x} \frac{p^r(1-p)^{n-r}\theta^{x-r}}{r!(x-r)!} .$$

By decomposing $(1-p)^{n-r}$ into $(1-p)^{n-x}(1-p)^{x-r}$ and multiplying by $x!/x!$, we obtain

(7c) $$f(x|p,n,\theta) = \frac{n!}{x!(n-x)!} (1-\theta)^{n-x}(1-p)^{n-x}x! \sum_{r=0}^{x} \frac{p^r[\theta(1-p)]^{x-r}}{r!(x-r)!}$$

$$= \frac{n!}{x!(n-x)!} [(1-\theta)(1-p)]^{n-x} \sum_{r=0}^{x} \frac{x!}{r!(x-r)!} p^r[\theta(1-p)]^{x-r}.$$

Since the summation term on the right is the binomial expansion of $[p+\theta(1-p)]^x$, we have

(7d) $$f(x|p,n,\theta) = \frac{n!}{x!(n-x)!} [p+\theta(1-p)]^x[(1-\theta)(1-p)]^{n-x}$$

which is, of course, an individual term of the binomial distribution with parameters, n and $p+\theta(1-p)$, i.e.,

(7e) $$f(x|p,n,\theta) = f_b(x|p+\theta(1-p), n).$$

Since

(8) $$p + \theta(1-p) = (1-\theta)p + \theta,$$

it should be clear that the existence of guessing (θ greater than 0) effects a linear transformation on the probability parameters of the non-guessing binomial distributions of r, given p.

As mentioned before, this result greatly simplifies the computations involved in obtaining the numerical results given later in this report. But, in addition, it has a more important implication. The existence of guessing under the conditions assumed in this basic model for testing does not change the form of any of the distributions of test statistics, since the basic conditional score distributions remain binomial. Therefore, without separate knowledge concerning either p or θ, it is impossible to detect or to isolate the effects of guessing using only the data available from the particular test administration.

The conditional distributions of x, given p, for $\theta = 1/5$, for $\theta = 1/2$ are given in Tables 5 and 6. The joint probability distributions of x and p are obtained as before and are given in Table 7 and 8. Though these joint distributions contain all of the information in the formal testing model, they fail to express a very important piece of information. What do we know about a person's ability level after we have observed his test score? This information is expressed by the conditional distributions of p, given x, which can be readily computed from the joint and marginal distributions given in Tables 2, 7, and 8. According to the basic definition of conditional probability, $P(B|A) = P(AB)/P(A)$. Thus, the conditional distribution of p for each x is obtained by dividing each joint probability by the appropriate marginal probability of x. These conditional distributions are given in Tables 9, 10, and 11. The marginal distributions of x for each of the three degrees of guessing, ($\theta = 0, 1/5, 1/2$) are shown in Figure 2 while the conditional distributions of p, given x are shown in Figure 3. Notice that increasing the degree of guessing makes the originally symmetric score distribution become negatively skewed. Observe also that increased guessing moves the conditional distributions of x, given p, away from the extremes of 0 and 1 and increases the spread of these distributions. This means that less information is being obtained concerning the actual ability level of the pupil. This is one way of expressing the degrading effects of guessing upon test information. Now we turn to a quantitative analysis of the effect of guessing upon decisions based upon this test information.

SELECTION, CLASSIFICATION AND PLACEMENT DECISIONS

The selection problem typically encountered in testing applications uses a test score, c, often called a cutting score to divide tested individuals into two groups. Those individuals with a test score of c or above are of further concern to the institution, since these individuals are chosen to have further interaction with the institution. For example, they are admitted into college, they are given

Table 5.

Conditional distributions of x given p. Minimal guessing (θ = 1/5).
Entries to be scaled times 10^{-5}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10		4	27	145	605	2114	6[illegible]29	17490	43439
9	8	65	346	1334	4031	9948	20302	33315	37773
8	88	520	1983	5543	12093	21066	28850	28555	14780
7	604	2465	6728	13643	21499	26436	24294	14504	3428
6	2720	7669	14986	22040	25082	21770	13426	4835	521
5	8392	16361	22888	24413	20066	12295	5068	1105	55
4	17982	24239	24275	18779	11148	4821	1339	175	4
3	26423	24623	17655	9906	4247	1296	241	20	
2	25479	16416	8426	3429	1061	229	29	1	
1	14560	6485	2383	703	158	24	2		
0	3744	1153	303	65	10	1			

Table 6.

Conditional distributions of x given p. Maximal guessing (θ = 1/2).
Entries to be scaled times 10^{-5}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10	253	605	1347	2824	5632	10738	19688	34868	59874
9	2072	4029	7249	12106	18771	26843	34742	38742	31512
8	7630	12093	17565	23347	28157	30199	27590	19371	7463
7	16648	21500	25222	26683	25028	20133	12983	5739	1048
6	23838	25082	23767	20012	14599	8808	4010	1116	97
5	23403	20066	15357	10292	5840	2642	849	149	6
4	15957	11148	6891	3676	1622	551	125	14	
3	7460	4247	2120	900	309	79	12	1	
2	2289	1062	428	145	39	7	1		
1	416	157	51	14	3				
0	34	11	3	1					

Table 7.

Joint and marginal distributions of x on p. Minimal guessing (θ = 1/5). Entries to be scaled times 10^{-5}.

Score (x)	Ability Level (p)									Total P(x)
	.1	.2	.3	.4	.5	.6	.7	.8	.9	
10			3	28	132	409	833	1003	443	2851
9		4	45	258	882	1925	2632	1910	385	8041
8	1	30	257	1072	2645	4076	3740	1637	151	13609
7	6	141	872	2640	4702	5115	3149	832	35	17492
6	28	440	1943	4264	5485	4212	1740	277	5	18394
5	86	938	2967	4724	4388	2379	660	63	1	16206
4	183	1389	3147	3634	2438	933	174	10		11908
3	270	1412	2289	1917	929	251	31	1		7100
2	260	941	1092	663	232	44	4			3236
1	148	372	309	136	35	5				1005
0	38	66	39	13	2					158
Total P(p)	1020	5733	12963	19349	21870	19349	12963	5733	1020	100000

Table 8.

Joint and marginal distributions of x and p. Maximal guessing (θ = 1/2). Entries to be scaled times 10^{-5}.

Score (x)	Ability Level (p)									Total P(x)
	.1	.2	.3	.4	.5	.6	.7	.8	.9	
10	3	35	175	547	1232	2078	2552	1999	611	9232
9	21	231	940	2342	4105	5194	4504	2221	321	198
8	78	693	2277	4518	6158	5843	3576	1110	76	24329
7	170	1233	3269	5163	5474	3896	1683	329	11	21228
6	243	1438	3081	3872	3193	1704	520	64	1	14116
5	239	1150	1991	1991	1277	511	110	9		7278
4	163	638	893	711	355	107	16	1		2884
3	76	244	275	174	67	15	2			853
2	23	61	55	28	8	1				176
1	4	9	7	3	1					24
0		1								1
Total P(p)	1020	5733	12963	19349	21870	19349	12963	5733	1020	100000

Table 9.

Conditional distributions of p given x. No guessing ($\theta = 0$).
Entries to be scaled times 10^{-3}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10				1	15	79	248	416	241
9				7	47	172	347	340	87
8			2	24	114	276	357	204	23
7			9	67	208	337	280	94	5
6		2	31	143	297	321	172	33	1
5		9	83	241	354	241	83	9	
4	1	33	172	321	297	143	31	2	
3	5	94	280	337	208	67	9		
2	23	204	357	276	114	24	2		
1	87	340	347	172	47	7			
0	241	416	248	79	14	1			

Table 10.

Conditional distributions of p given x. Minimal guessing ($\theta = 1/5$).
Entries to be scaled times 10^{-3}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10			1	10	46	144	292	352	155
9		1	6	32	110	239	327	237	48
8		2	19	79	194	300	275	120	11
7		8	50	151	269	292	180	48	2
6	1	24	106	232	298	229	95	15	
5	5	58	183	291	271	147	41	4	
4	15	117	264	305	205	78	15	1	
3	38	199	322	270	131	36	4		
2	80	291	337	205	72	14	1		
1	148	370	307	135	35	5			
0	240	418	247	82	13				

Table 11.

Conditional distributions of p given x. Maximal guessing ($\theta = 1/2$).
Entries to be scaled times 10^{-3}.

Score (x)	Ability Level (p) .1	.2	.3	.4	.5	.6	.7	.8	.9
10		4	19	59	133	225	277	217	66
9	1	12	47	118	206	261	227	112	16
8	3	28	94	186	253	240	147	46	3
7	8	58	154	243	258	184	79	16	
6	17	102	218	274	226	121	37	5	
5	33	158	274	274	175	70	15	1	
4	56	221	310	247	123	37	6		
3	89	286	322	204	79	18	2		
2	131	342	312	158	48	8	1		
1	183	387	285	117	28				
0	224	404	250	122					

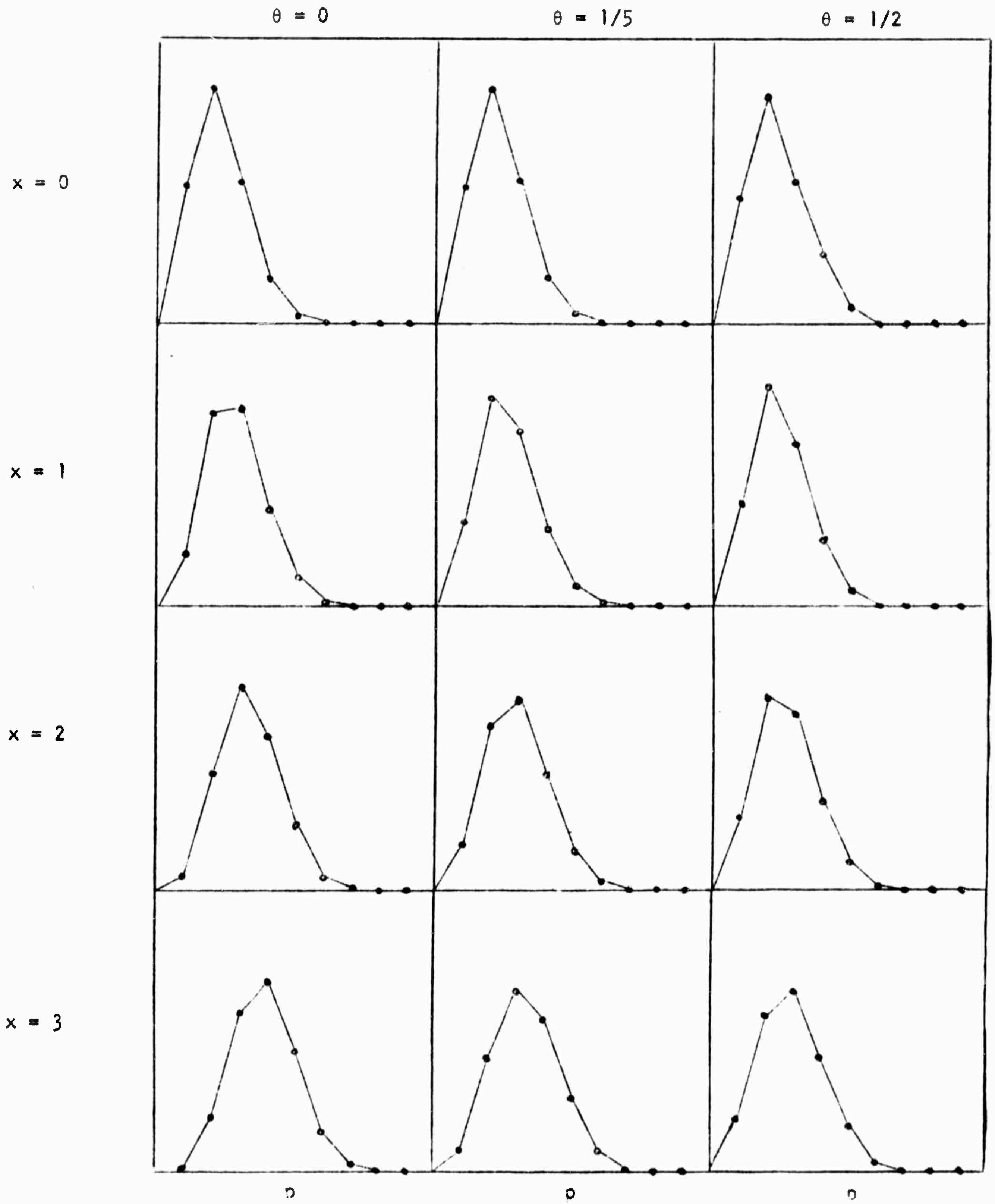


Figure 3a. Conditional distributions of p given x for a 10-item test affected by different degrees of guessing.

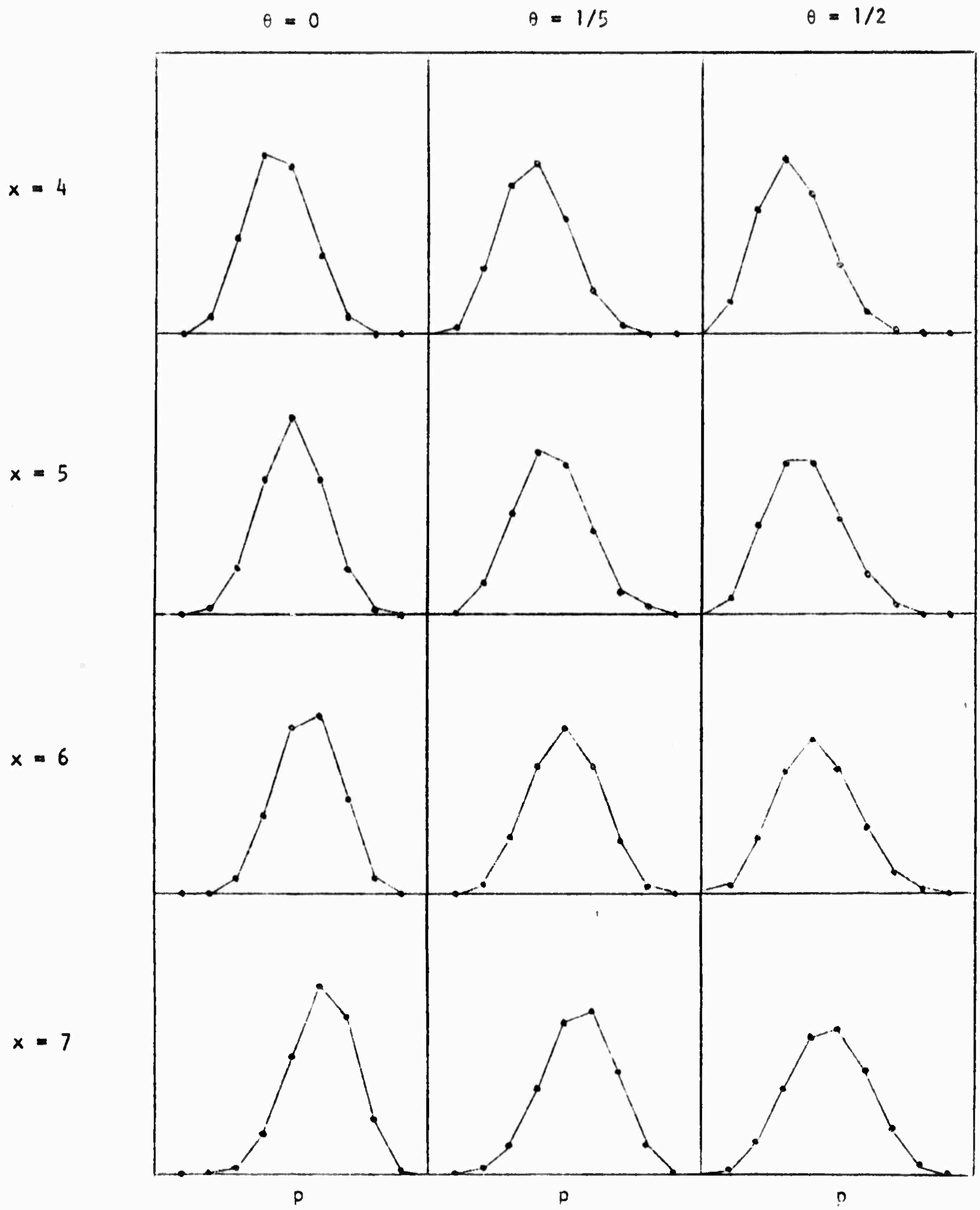


Figure 3b. Conditional distributions of p given x for a 10-item test affected by different degrees of guessing.

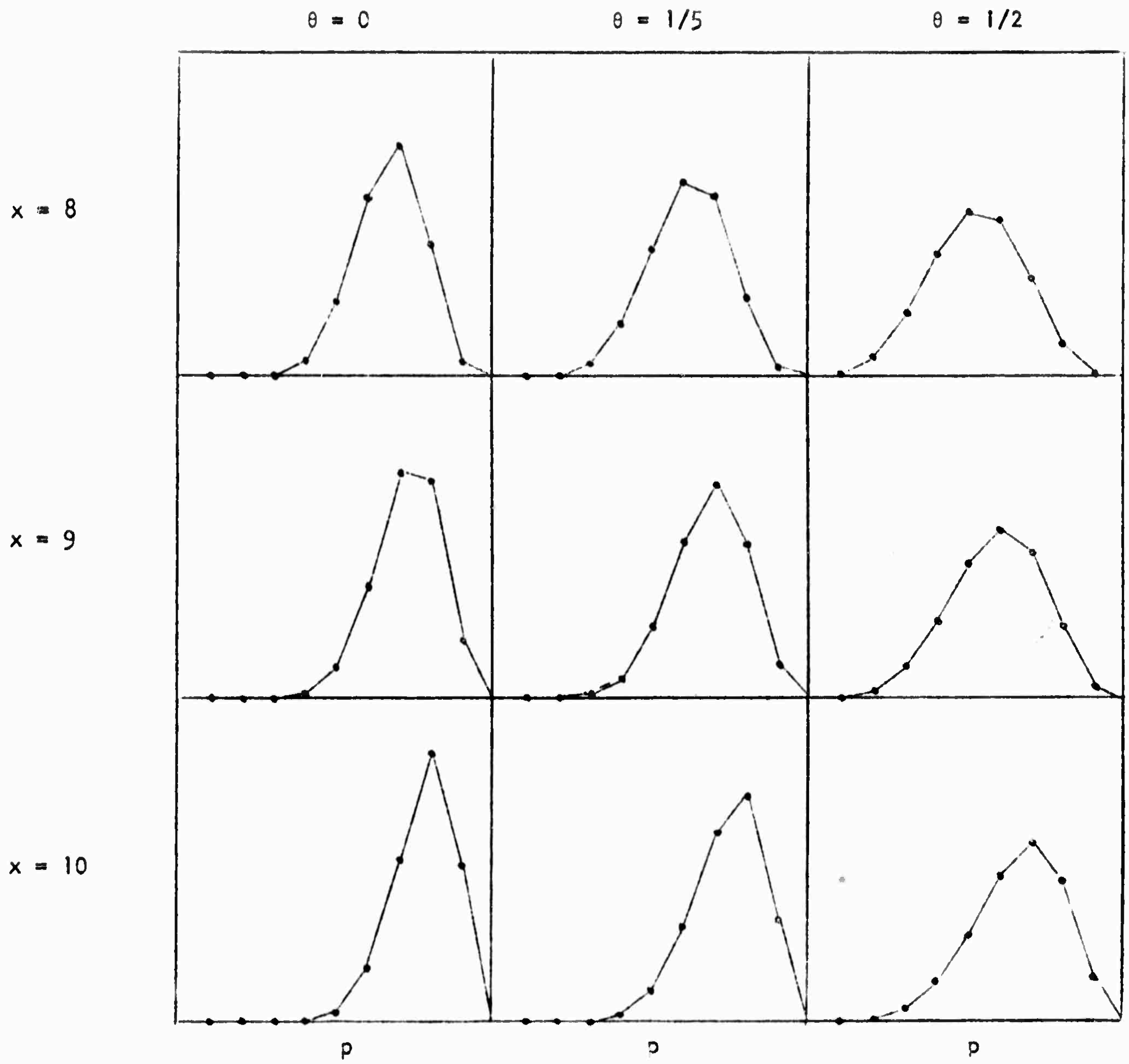


Figure 3c. Conditional distributions of p given x for a 10-item test affected by different degrees of guessing.

flying training, or they may be employed by a company. The value or utility to the institution of one of these chosen individuals is often approximated by a linear function of ability level, p, which may be written as

(9) $$U(p) = kp + K, \; k > 0, \; K < 0$$

where k must be greater than zero in order to keep the problem from becoming trivial and to imply that the institution desires people with high ability levels. Those individuals (scoring $x < c$) not chosen by the institution are of no further concern to the organization; thus the value or utility to the institution of not selecting an individual is usually taken to be zero.

As emphasized repeatedly by Cronbach and Gleser (1965) the performance of any testing process should not be compared with some chance level, but should be compared with how well the process could be effected by taking into account all the information available from sources other than testing. Within our formal model all information of this type is expressed by the marginal probability distribution of p, P(p). Thus, we began by computing the expected return from a selection process based not upon testing but upon all other available information. But first, it is convenient to rewrite the parameters of the utility function. Let p_o be that ability level that yields a return of zero to the institution. This allows us to express K in terms of k and p_o, that is

(10) $$K = -kp_o$$

and the utility function can now be written as

(11) $$U(p) = k(p-p_o).$$

The value to the institution of selecting an individual must take account of the uncertainty about the individual's true ability level. Thus, the expected value of selection using no testing information is

(12) $$E'U(p) = \sum_p k(p-p_o)P(p)$$
$$= k(\bar{p}'-p_o)$$

where $\bar{p}'$ is the mean of the prior or initial distribution, of p. Notice that if this average ability level is less than p_o, Equation 12 becomes negative implying that on the average the institution loses by selecting individuals. In this case, no individual should be selected, yielding a zero return to the institution, which is not good, but it is clearly better than a negative return. In order to compare the gain due to selection testing, the largest of these two values, either zero or the expected value of selection, must be subtracted from the expected utility achieved by selection testing.

Now consider the expected value of selection testing where individuals are selected or rejected on the basis of their test score, x. If an individual earns a test score, x, the expected value of selection is

(13)
$$\begin{aligned} E''_x U(p) &= \sum_p k(p-p_o)P(p|x) \\ &= k(\bar{p}''_x - p_o) \end{aligned}$$

where $\bar{p}''_x$ is the mean of the conditional distribution of p, given x or, analogously, the average ability level of those individuals making a test score of x. Observe that selection is of value to the institution whenever the selected individual's test score implies an average ability level greater than p_o. Now, consider setting a cutting score, c, so that all individuals with scores of c or above are selected and all others rejected. The expected value to the institution of such a decision rule must be computed by taking account of the frequency with which individuals will obtain the different test scores and can be expressed as

(14)
$$\begin{aligned} E''_c U(p) &= \sum_{x=c}^{n} k(\bar{p}''_x - p_o)P(x) \\ &= k\{\sum_{x=c}^{n} \bar{p}''_x P(x) - p_o \sum_{x=c}^{n} P(x)\} \end{aligned}$$

The expected value of selection testing with a cutting score, c, can vary over a wide range depending upon the choice of the cutting score. The optimal decision rule is obtained by selecting that cutting score, c*, which yields the largest expected value for selection testing. Notice that the selection ratio is not explicitly taken into account here, though the last term on the right in (14) incorporates the selection ratio. Therefore, selecting the best cutting score, c*, also fixes the corresponding selection ratio.

To obtain the expected value to the institution, of selection testing, we must subtract the expected value of the selection process, not using testing information, from the expected value of selection testing. Thus,

$$(15)\qquad EVST = \begin{cases} k\{\sum_{x=c}^{n}(\bar{p}_x''-p_o)P(x)-(\bar{p}'-p_o)\} & \text{if } E'U(p)>0 \\ k\sum_{x=c}^{n}(\bar{p}_x''-p_o)P(x) & \text{if } E'U(p)\leq 0 \end{cases}$$

Notice that the advantage of rewriting the utility function now becomes apparent. All terms are now multiplied by the slope constant, k, which means that computations can be performed leaving k as an unspecified parameter. Therefore, in considering any practical decision problem, all we need to do is to specify k and p_o in order to obtain absolute utility values appropriate to the problem. Table 12 gives the expected value of selection testing for different cutting scores, three levels of p_o and for the three levels of guessing. The entries enclosed by rectangles correspond to the maximum return possible and identify the optimal cutting score, c*.

Figure 3 graphs these values to illustrate the effects of guessing. Notice that the effect of guessing is both to increase the optimal cutting score and to decrease the expected value of selection testing with this cutting score. Notice also that the choice of cutting score can be quite critical particularly when p_o is not equal to 1/2, in this case, the average ability level for the population. Especially notice that the expected value of selection testing can become quite negative which may represent a considerable loss to the institution. Clearly, the specification of a program for selection testing is not to be undertaken lightly and the higher-level institutional decision to adopt selection testing should be based on firm assurances that optimal cutting scores have been adopted which do not represent a loss to the institution. Finally, observe that the cost of testing is independent of the cutting score. Thus, the cost of testing divided by k can be plotted on these graphs as a horizontal line with some positive height above zero. In effect, this serves to move the zero point of the scale along the ordinate to some higher point corresponding to the cost of testing divided by k. It should be clear that this could serve to reduce the number of situations in which testing has any positive value. Notice, in particular, the graph for maximum guessing, $\theta = 1/2$. If the cost of the testing program were at all significant, it could easily exceed the rather small returns of selection testing when p_o is equal to .4 or to .6. Another comparison of some significance can be made. So far, we have considered the added value of testing relative to not testing in a selection process.

Table 12.

Expected value of selection testing

Guessing Probability	Cutting Score (c)	$\sum_{x \geq c}^{n} \bar{p}''_x P(x)$	$\sum_{x \geq c}^{n} P(x)$	Cutting Score (c)	$\frac{1}{k}$, Value of Selection Testing $p_o = .4$	$p_o = .5$	$p_o = .6$
	11	0	0	11	-10000	0	0
	10	1151	1478	10	- 9440	412	264
	9	4423	6008	9	- 7980	1419	818
	8	10077	14488	8	- 5718	2833	1384
	7	17616	26826	7	- 3114	4203	$\boxed{1520}$
	6	26013	41940	6	- 763	5043	849
$\theta = 0$	5	34074	58060	5	850	$\boxed{5044}$	- 762
	4	40791	73174	4	$\boxed{1521}$	4204	- 3113
	3	45589	85512	3	1384	2833	- 5718
	2	48416	93992	2	819	1420	- 7979
	1	49673	98522	1	264	412	- 9440
	0	50000	100000	0	0	0	-10000
	11	0	0	11	-10000	0	0
	10	2108	2851	10	- 9032	682	397
	9	7538	10892	9	- 6829	2092	1003
	8	15882	24501	8	- 3918	3632	$\boxed{1181}$
	7	25550	41993	7	- 1247	$\boxed{4554}$	354
$\theta = \frac{1}{5}$	6	34645	60387	6	490	4452	- 1587
	5	41756	76593	5	$\boxed{1112}$	3460	- 4200
	4	46357	88501	4	957	2106	- 6744
	3	48758	95601	3	518	958	- 8603
	2	49710	98837	2	175	292	- 9592
	1	49967	99842	1	30	46	- 9938
	0	50000	100000	0	0	0	-10000
	11	0	0	11	-10000	0	0
	10	6076	9232	10	- 7617	1460	$\boxed{537}$
	9	17731	29112	9	- 3914	3175	264
	8	30413	53441	8	- 963	$\boxed{3692}$	- 1652
	7	40248	74669	7	380	2914	- 4553
$\theta = \frac{1}{2}$	6	46068	88785	6	$\boxed{554}$	1676	- 7203
	5	48744	96063	5	319	712	- 8894
	4	49694	98947	4	115	220	- 9674
	3	49947	99800	3	27	47	- 9933
	2	49995	99976	2	5	7	- 9991
	1	50000	100000	1	0	0	-10000
	0	50000	100000	0	0	0	-10000

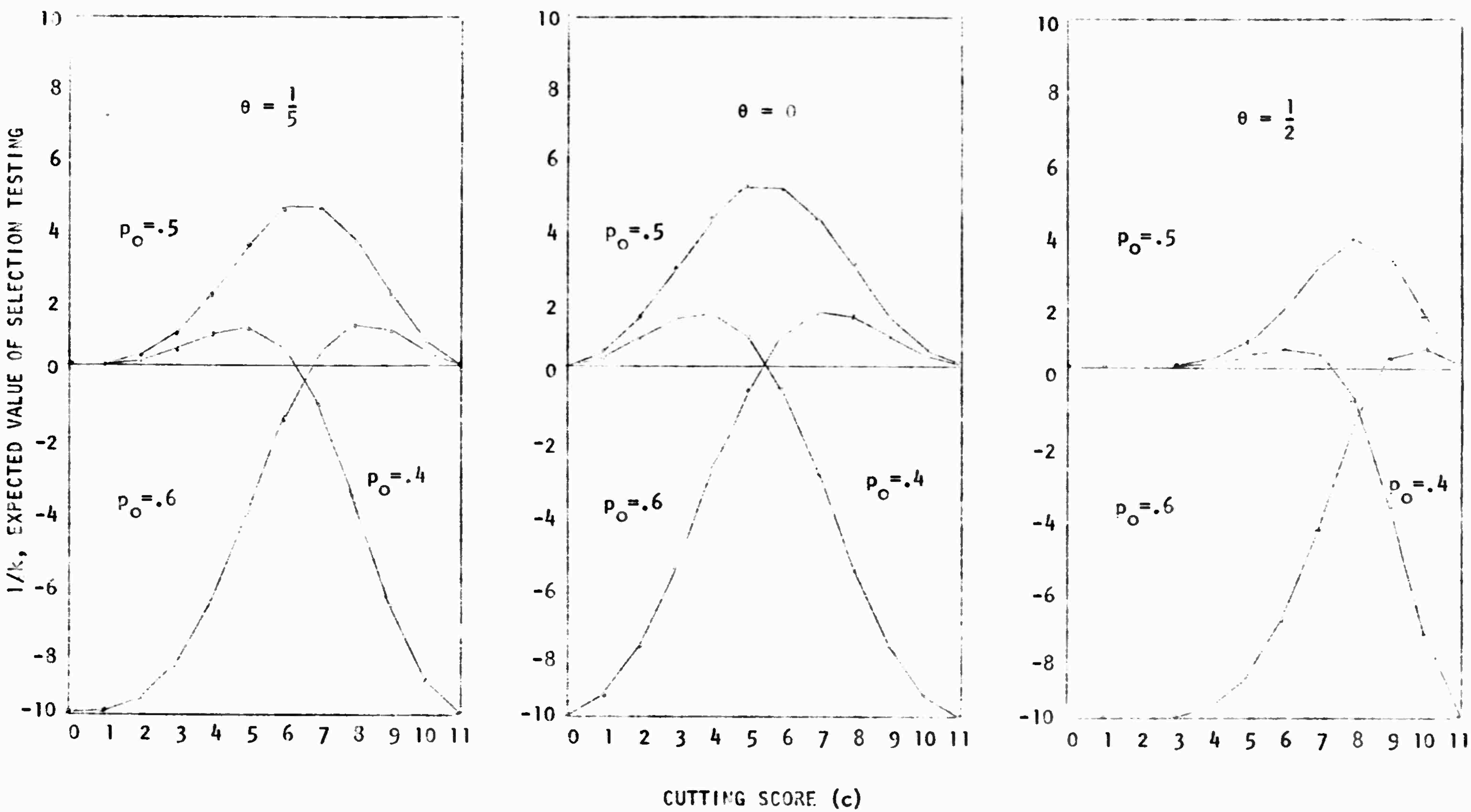


Figure 3. Behavior of expected value of selection testing, × (1/k) for different levels of guessing.

What would be the added value of knowing exactly each individual's true ability level? This could be known, in principle, if we used an admissible test, which eliminated guessing and used all the items in the pool. Let us define the expected value of perfect information as the gain in expected value to the institution resulting from having perfect knowledge of each individual's ability level relative to that of having imperfect non-testing information as to an individual's ability level. Thus, we have

$$\text{(16)} \qquad \text{EVPI} = \begin{cases} k\{\sum\limits_{p \geq p_c} (p-p_o)P(p)-(\bar{p}'-p_o)\} & \text{if } \bar{p}' > p_o \\ k\sum\limits_{p \geq p_c} (p-p_o)P(p) & \text{if } \bar{p}' \leq p_o \end{cases}$$

Equation 16 can be used both to find the optimal cutting point along ability level and the corresponding expected value to the institution. Table 13 shows these optimal cutting points and expected values for various critical ability levels, p_o. Table 13 also shows optimal cutting scores and expected values for our 10-item test affected by the three different degrees of guessing. Notice that for extreme critical ability levels, even perfect information does not help. The expected values are zero and one can do as well by accepting all individuals in the case of very high critical ability levels or rejecting all individuals in the case of very low critical ability levels. If one considers the 10-item test, the range of critical ability level for which testing yields a gain is narrowed even more. It is, of course, narrowed further by the existence of higher degrees of guessing. These and other relations are grasped more easily by examining Figure 4. Notice first that these functions are symmetric and p_o equal to 1/2, which, remember, corresponds to the average ability level of the population of individuals. If the average ability level were some other value, then these functions would be shifted to either the right or the left. Information concerning an individual's ability level is of most value when the critical ability level, p_c, of the utility function is near the average ability level, $\bar{p}'$. The value of this information falls off quite rapidly to either side of average ability level and declines down to a value of zero corresponding to the value of the selection process without the use of additional information. This is a reflection of the generalization that additional information cannot hurt. It is true, however, only because these are optimal selection processes based upon the best cutting score. Use of any but this one best cutting score could easily represent a significant loss to the institution.

Notice the vertical distance between the function showing the expected gain

Table 13.

Optimal selection with perfect information and with a 10-item test affected by different degrees of guessing.

Critical Ability Level	Perfect Information		No Guessing $\theta = 0$		Minimum Guessing $\theta = 1/5$		Maximum Guessing $\theta = 1/2$	
(p_o)	p_c^*	$\frac{1}{k}$EVPI	c^*	$\frac{1}{k}$EVT	c^*	$\frac{1}{k}$EVT	c^*	$\frac{1}{k}$EVT
1.00	1.0	0	11	0	11	0	11	0
.95	1.0	0	11	0	11	0	11	0
.90	.9	0	11	0	11	0	11	0
.85	.9	51	11	0	11	0	11	0
.80	.8	102	11	0	11	0	11	0
.75	.8	439	10	43	11	0	11	0
.70	.7	777	9	217	10	112	11	0
.65	.7	1763	8	660	9	458	10	75
.60	.6	2749	7	1520	8	1181	10	537
.55	.6	4702	6	2946	7	2454	9	1719
.50	.5	6656	5	5044	7	4554	8	3692
.45	.5	4702	5	2947	6	2471	7	1647
.40	.4	2749	4	1521	5	1112	6	554
.35	.4	1763	3	660	4	382	5	122
.30	.3	777	2	218	3	78	4	10
.25	.3	439	1	42	1	6	2	1
.20	.2	102	0	0	0	0	0	0
.15	.2	51	0	0	0	0	0	0
.10	.1	0	0	0	0	0	0	0
.05	.1	0	0	0	0	0	0	0
.00	0	0	0	0	0	0	0	0

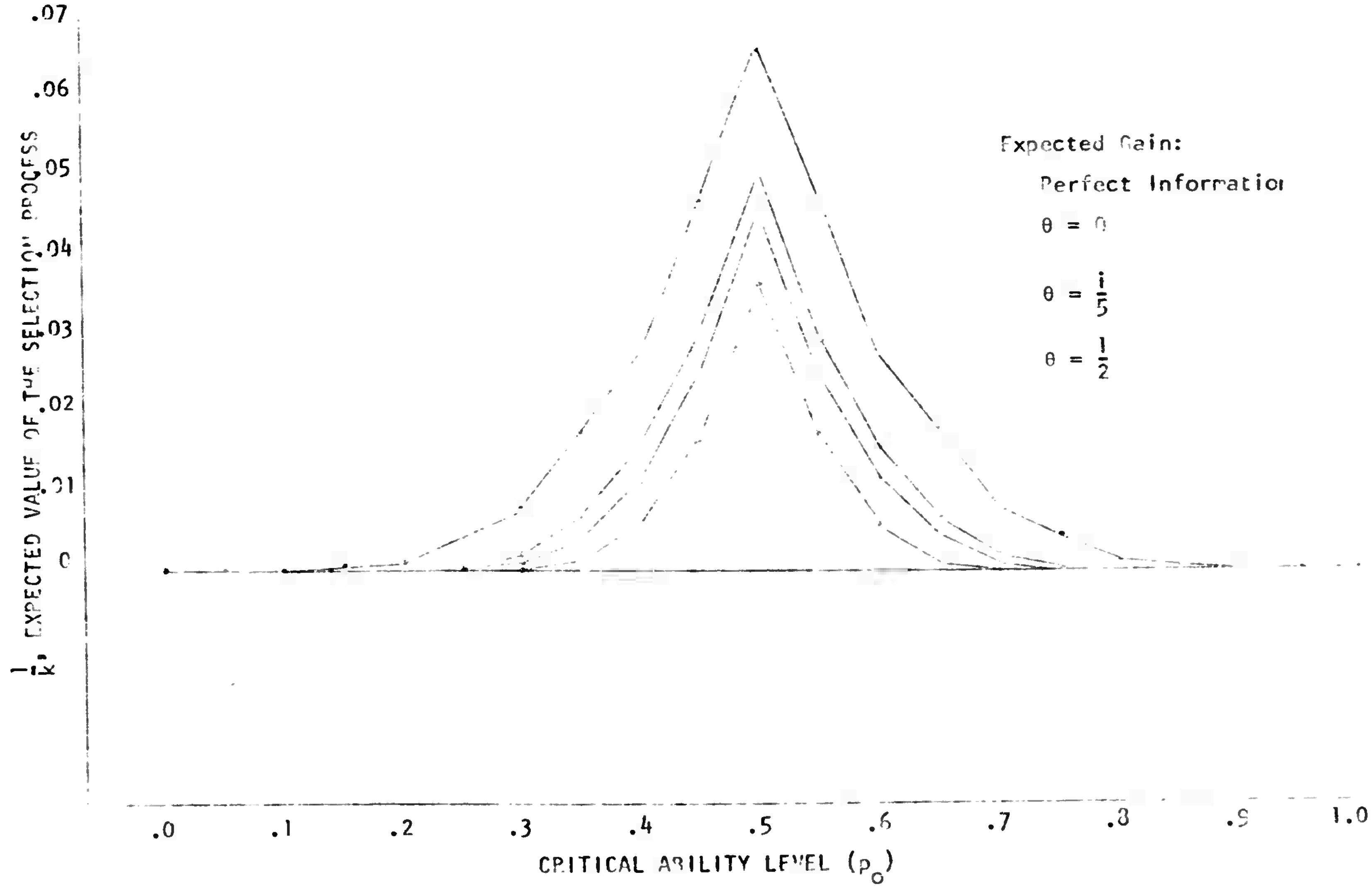


Figure 4. Relative expected value of optimal selection processes.

for perfect information and that showing the expected gain for a 10-item test with no guessing. This distance represents, in a sense, the loss due to sampling of test items or, conversely, the maximum gain possible by lengthening the test without bringing in any guessing behavior. Now notice the vertical distance between the function for a ten-item test with no guessing and the function for a ten-item test with maximum guessing ($\theta = 1/2$). This distance represents the maximum possible loss due to guessing or, conversely, the maximum possible gain due to the elimination of guessing, but keeping the same test length. Observe that these two sets of distances are approximately the same size which implies that the elimination of guessing on a ten-item test could yield benefits comparable to those obtainable by changing from a ten-item test free of guessing to a test of nearly infinite length which is also free of guessing. In this sense, the deterioration in performance of selection testing which may be attributed to the effect of guessing is enormous.

As in the case of Figure 3, the effect of adding in the cost of a testing program can be represented by a horizontal line placed at that value of the ordinate corresponding to the cost of the testing program. This, in effect, raises the zero point along the scale of the ordinate and implies, that adoption of a selection testing program by an institution when the critical ability level is extreme represents a gross loss to the institution. The added cost of modifying testing procedures so as to eliminate guessing should be a small fraction of the present cost of operating a testing program which is composed largely of administration costs. Therefore, over the range of situations for which a selection testing program is of benefit to the institution, the net gain of eliminating guessing will be of a quite appreciable magnitude.

Consider now a placement process, where individuals are assigned to one of two programs. These programs may represent different instructional methods or classes grouped according to ability level, two different schools, two different jobs, or two different psychiatric treatments. (See Cronbach and Gleser, 1965). The utility to the institution of assigning an individual to either of the two programs is assumed to be a linear function of the individual's ability level and may be written as

$$(9') \qquad U_1(p) = k_1p+K_1, \quad U_2(p) = k_2p+K_2, \quad k_1>k_2, \; K_1< K_2$$

Since we are interested in the relative performance of various placement processes it is convenient to rewrite the utility functions as gain functions. Thus, (9'a) $$G_1(p) = (k_1-k_2)p+K_1-K_2, \quad G_2(p) = (k_2-k_1)p+K_2-K_1$$

The original utility functions and the revised gain functions are shown in Figure 5. The break-even point, p_b, where the functions intersect may be obtained by setting $G = 0$, thus

(10') $$p_b = \frac{K_2-K_1}{k_1-k_2} = \frac{K_1-K_2}{k_2-k_1}$$

and the gain functions may be rewritten as

(11') $$G_1(p) = (k_1-k_2)(p-p_b), \quad G_2(p) = (k_1-k_2)(p_b-p)$$

Given only the non-testing information expressed in the marginal distribution of p the expected gains from placing an individual in Program 1 or Program 2 may be computed as

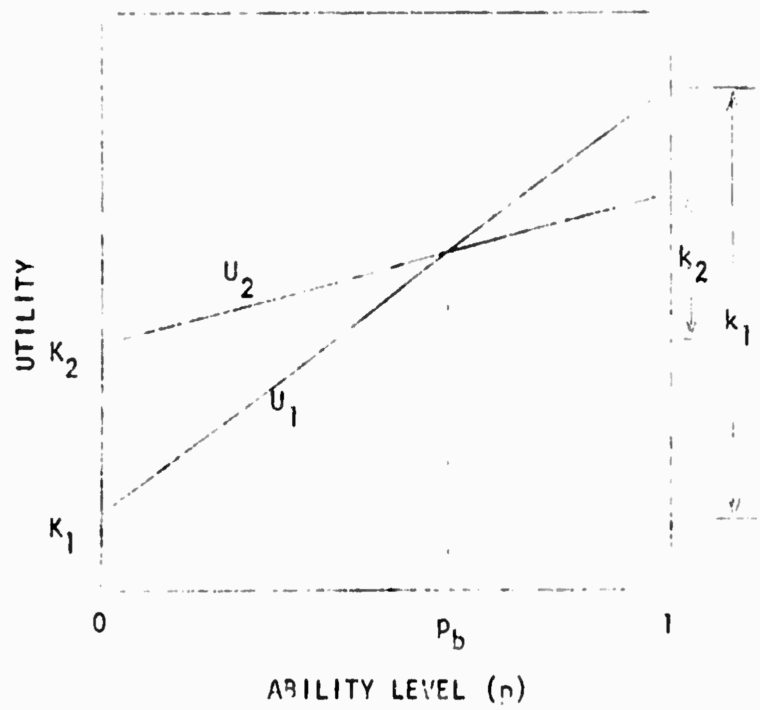
(12') $$E'G_1(p) = \sum_p (k_1-k_2)(p-p_b)P(p), \quad E'G_2(p) = \sum_p (k_1-k_2)(p_b-p)P(p)$$
$$= (k_1-k_2)(\bar{p}'-p_b) \qquad\qquad = (k_1-k_2)(p_b-\bar{p}')$$

Notice that the factor k_1-k_2 must be positive. Therefore, the expected gain will be positive or negative depending upon whether the second factor is positive or negative. It will be positive if average ability level, $\bar{p}'$, is greater than the break-even point, p_b. In this case, the individual should be assigned to Program 1. However, if the average ability level is smaller than the break-even point, the expected gain for Program 2 will be larger and the individual should be assigned to Program 2.

In placement testing, the expected gain depends upon an individual's test score, x, and may be written as

(13') $$E''_x G_1(p) = \sum_p (k_1-k_2)(p-p_b)P(p|x) = (k_1-k_2)(\bar{p}''_x-p_b)$$
$$E''_x G_2(p) = \sum_p (k_1-k_2)(p_b-p)P(p|x) = (k_1-k_2)(p_b-\bar{p}''_x)$$

The overall expected gain from placement testing is a weighted sum of the



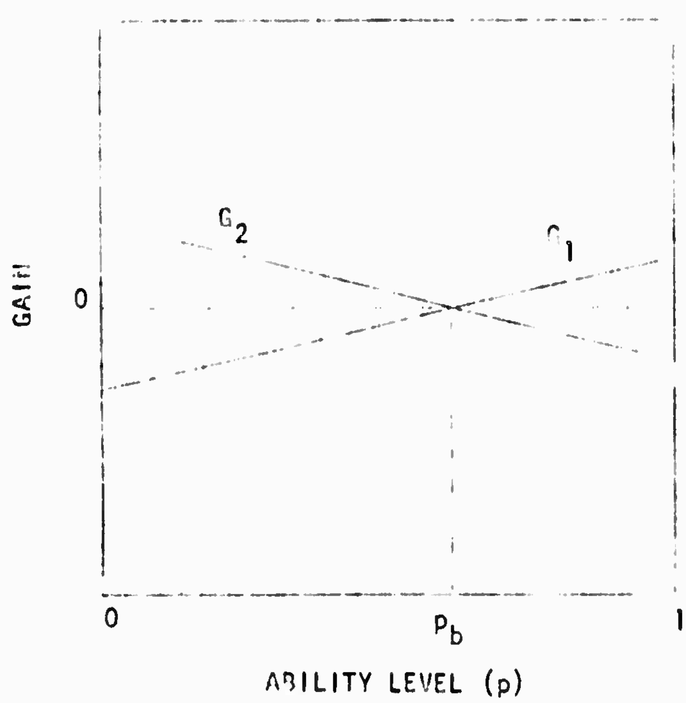


Figure 5. Utility and gain functions for a placement decisions.

conditional gains, thus

(14')
$$E_c''G(p) = (k_1-k_2)\{\sum_{x=c}^{n} (\bar{p}_x''-p_b)P(x)+\sum_{x=0}^{c-1} (p_b-\bar{p}_x'')P(x)\}$$
$$= (k_1-k_2)\{\sum_{x=c}^{n} \bar{p}_x''P(x)-\sum_{x=0}^{c-1} \bar{p}_x''P(x)+p_b[\sum_{x=0}^{c-1} P(x)-\sum_{x=c}^{n} P(x)]\}$$

The expected gain from a placement process using only non-testing information must be subtracted from this value to obtain the expected value of placement testing. Thus,

(15')
$$\text{EVPT} = \begin{cases} E_c''G(p)-(k_1-k_2)(\bar{p}'-p_b) & \text{if } \bar{p}' > p_b \\ E_c''G(p)-(k_1-k_2)(p_b-\bar{p}') & \text{if } \bar{p}' < p_b \end{cases}$$

The expected value of perfect information is

(16')
$$\text{EVPI} = \begin{cases} (k_1-k_2)\{\sum_{p\geq p_c} pP(p)-\sum_{p<p_c} pP(p)+p_b[\sum_{p<p_c} P(p)-\sum_{p\geq p_c} P(p)]-(\bar{p}'-p_b)\} & \text{if } \bar{p}' > p_b \\ (k_1-k_2)\{\sum_{p\geq p_c} pP(p)-\sum_{p<p_c} pP(p)+p_b[\sum_{p<p_c} P(p)-\sum_{p\geq p_c} P(p)]-(p_b-\bar{p}')\} & \text{if } \bar{p}' < p_b \end{cases}$$

The expected value of placement testing has been computed and is shown in Table 14 for three levels of guessing and for three break-even points. Notice that the optimal cutting scores are the same as those shown in Table 12. Notice further that the expected values are twice those shown in Table 12. This suggests a theorem.

THEOREM 2. If $p_o = p_b$,

(17)
$$\frac{\text{EVPT}}{k_1-k_2} = 2\,\frac{\text{EVST}}{k}$$

Proof: We will prove this theorem by deriving the basic equations in a somewhat different manner from that above. In addition to enabling us to prove the theorem, this may have added heuristic value in understanding the basic relations. The theorem must be proved seperately for two cases.

Table 14.

Expected value of placement testing $(x[k_1-k_2]^{-1})$.

Guessing Probability θ	Cutting Score (c)	$\sum_{x=c}^{n} \bar{p}''_x P(x) - \sum_{x=0}^{c-1} \bar{p}''_x P(x)$	$\sum_{x=0}^{c-1} P(x) - \sum_{x=0}^{n} P(x)$	$\frac{1}{(k_1-k_2)}$ Expected Value of Testing $p_b = .4$	$p_b = .5$	$p_b = .6$
$\theta = 0$	11	-50000	100000	-20000	0	0
	10	-47698	97044	-19880	824	528
	9	-41154	87984	-11960	2838	1636
	8	-29846	71024	-11436	5666	2768
	7	-14768	46348	- 6229	8406	$\boxed{3041}$
	6	2026	16120	- 1526	10086	1698
	5	18148	-16120	1700	$\boxed{10086}$	- 1524
	4	31582	-46348	$\boxed{3043}$	8408	- 6227
	3	41178	-71024	2768	5666	-11436
	2	46832	-87984	1638	2840	-15958
	1	49346	-97044	528	824	-18880
	0	50000	-100000	0	0	-20000
$\theta = \frac{1}{5}$	11	-50000	100000	-20000	0	0
	10	-45786	94298	-18067	1363	793
	9	-34925	78216	-13639	4183	2005
	8	-18237	50998	- 7838	7262	$\boxed{2362}$
	7	1098	16014	- 2496	$\boxed{9105}$	706
	6	19289	-20774	979	8902	- 3175
	5	33510	-53186	$\boxed{2236}$	6917	- 8402
	4	42712	-77002	1911	4211	-13489
	3	47514	-91202	1033	1913	-17207
	2	49418	-97679	346	578	-19189
	1	49932	-99684	58	90	-19878
	0	50000	-100000	0	0	-20000
$\theta = \frac{1}{2}$	11	-50000	100000	-20000	0	0
	10	-37848	81536	-15234	2920	$\boxed{1074}$
	9	-14539	41776	- 7829	6349	527
	8	10825	- 6882	- 1928	$\boxed{7384}$	- 3304
	7	30495	-49338	760	5826	- 9108
	6	42135	-77570	$\boxed{1107}$	3350	-14407
	5	47488	-92126	638	1425	-17783
	4	49388	-97894	230	441	-19348
	3	49893	-99600	53	93	-19867
	2	49989	-99952	8	13	-19982
	1	50000	-100000	0	0	-20000
	0	50000	-100000	0	0	-20000

Case I. Suppose that the average ability level is greater than the break-even point, i.e.,

(17a) $$\bar{p}' > p_b$$

This defines Case I.

Since the average ability level is greater than the break-even point, the optimal strategy for selection is to admit all individuals, thus

(17b) $$\frac{E'U(p)}{k} = \sum_{S} (p-p_b)P(x,p)$$

Notice that instead of taking the expectation by using the marginal probability of p, we are using the joint probability of x and p and summing over all possible combinations of x and p. This is the basic change in approach that we will use to prove this theorem.

In a similar fashion, the expected gain from selection testing may be written as

(17c) $$\frac{E'_cU(p)}{k} = \sum_{A} (p-p_b)P(x,p)$$

Here the summation over the set A includes all those pairs, (x,p), for which x is greater than or equal to c. Now, we can take the difference between the two previous equations to obtain the expected value of selection testing. Thus,

(17d) $$\begin{aligned}\frac{EVST}{k} &= \sum_{A} (p-p_b)P(x,p) - \sum_{S} (p-p_b)P(x,p)\\ &= -\sum_{B} (p-p_b)P(x,p)\\ &= \sum_{B} (p_b-p)P(x,p)\end{aligned}$$

Here, the summation over the set B includes all those pairs, (x,p), for which x is less than c.

For the placement decision, the expected value of placement using only non-testing information may be written as

(17e)
$$\frac{E'U(p)}{k_1-k_2} = \sum_S (p-p_b)P(x,p)$$

The expected gain for placement testing is

(17f)
$$\frac{E_c''G(p)}{k_1-k_2} = \sum_A (p-p_b)P(x,p) + \sum_B (p_b-p)P(x,p)$$

while the expected value of placement testing is

(17g)
$$\frac{EVPT}{k_1-k_2} = \sum_A (p-p_b)P(x,p) + \sum_B (p_b-p)P(x,p) - \sum_S (p-p_b)P(x,p)$$

$$= \sum_B (p_b-p)P(x,p) - \sum_B (p-p_b)P(x,p)$$

$$= 2\sum_B (p_b-p)P(x,p)$$

Now compare the last line in Equation 17g with the last line in Equation 17d. The ratio between placement and selection testing s 2:1 as was to be proved.

Case II.

$$\bar{p}' < p_b$$

In this case the average ability level is less than the break-even point. Therefore the optimal selection strategy under no test information is to reject every individual. Thus

(17'b)
$$\frac{E'U(p)}{k} = 0$$

By reasoning analogous to that used for Case I, we find that the expected value of selection testing may be written as

(17'd)
$$\frac{EVST}{k} = \sum_A (p-p_b)P(x,p)$$

As for placement testing, since the average ability level is less than the break-even point, the optimal strategy with no test information is to place all individuals in Program 2. Thus, (17'e) $$\frac{E'U(p)}{k_1-k_2} = \sum_S (p_b-p)P(x,p)$$

Again, reasoning as in Case I, we may write the expected value of placement testing as

(17'g) $$\begin{aligned}\frac{EVPT}{k_1-k_2} &= \sum_A (p-p_b)P(x,p) + \sum_B (p_b-p)P(x,p) - \sum_S (p_b-p)P(x,p)\\ &= \sum_A (p-p_b)P(x,p) - \sum_A (p_b-p)P(x,p)\\ &= 2\sum_A (p-p_b)P(x,p)\end{aligned}$$

As in Case I, compare the last line of Equation 17'g with Equation 17'd. The ratio between placement testing and selection testing is 2:1 as was to be proved. Q.E.D.

This theorem has a useful corollary.

Corollary: $(k_1-k_2)^{-1}$ times the expected value of perfect information for a placement decision is twice $(k)^{-1}$ times the expected value of perfect information for the corresponding selection decision.

Theorem 2 and its corollary imply that with a simple multiplicative adjustment all our results for selection testing hold also for placement testing, so the comparisons and comments made above apply with equal, if not greater, force to the placement decisions.

EDUCATIONAL AND VOCATIONAL COUNSELING DECISIONS

Information obtained from testing is frequently used to guide educational and vocational counseling decisions. Generally, a person's test score is used to estimate his ability level. This estimate is made part of his record and is then used over a period of time to guide both institutional and individual decisions. The essential characteristic of this class of applications is that an all-purpose estimate is obtained to be incorporated into many decision problems. In this sense, the use of testing information in counseling decisions is similar to the general problem of estimation of parameters in science. No attempt is made to tailor the estimate to one particular application, but the estimate is meant to serve many different applications.

The obtaining of such a general purpose estimate can itself be considered a decision problem with the different alternatives being the various possible estimated ability levels and the utility (here the distinction between individual and institutional decisions becomes blurred) being some function of the difference between estimated ability level and true ability level. The utility function most frequently used in this type of application is proportional to the complement of squared error. Thus

$$(18) \qquad U(\tilde{p},p) = k[1-(\tilde{p}-p)^2], \quad k > 0.$$

The use of the squared-error criterion means that the value of an estimation process will depend upon the variances of the distributions involved. For example, without the use of testing information, a person's ability level may be estimated by the mean of the marginal distribution of p. Thus

$$(19) \qquad \begin{aligned} EU(\bar{p}',p) &= \sum_p k[1-(\bar{p}'-p)^2]P(p) \\ &= k-k\sum_p (p-\bar{p}')^2P(p) \\ &= k[1-V'(p)] \end{aligned}$$

The second term on the right is the variance of the marginal distribution of p since the squared deviations are taken with respect to the mean of this distribution. It is, of course, well-known that the weighted sum of the squared deviations about the mean of a distribution is a minimum and that this mean is, thus, the best possible estimate for the squared error criterion.

Given the availability of testing information a mean still remains the best estimate, but the mean is conditional upon the test score, x, and the expectation is taken with respect to the conditional distribution of p given x. Thus,

$$(20) \qquad \begin{aligned} E_xU(\bar{p}''_x,p) &= k[1-\sum_p(p-\bar{p}''_x)^2P(p|x)] \\ &= k[1-V''_x(p)] \end{aligned}$$

where the second term in the brackets on the right is, of course, the variance of the conditional distribution of p.

The overall performance of the estimation process is obtained by taking the weighted sum of these conditional expected utilities, the weights being the marginal probabilities of x. Thus

$$(21) \qquad EE_x U(\bar{p}''_x,p) = \sum_{x=0}^{n} k[1-V''_x(p)]P(x)$$
$$= k[1-\sum_{x=0}^{n} V''_x(p)P(x)]$$
$$= k[1-V''(p)]$$

where the second term in brackets on the right is the expected value of the conditional variance.

Now, in terms of conventional test theory, V'p is the variance of the true scores while V''p is the variance remaining after testing. Thus, the difference between these two variances, V'p-V''p, is the variance accounted for by testing and the ratio of this difference to the initial variance, [V'p-V''p]/V'p, is an important measure of test performance. Applying these operations to our expected utility equations, we obtain

$$(22) \qquad \frac{EU(\bar{p}',p)-EE_x U(\bar{p}''_x,p)}{EU(\bar{p}',p)} = \frac{k[1-V'(p)]-k[1-V''(p)]}{k[1-V'(p)]}$$
$$= \frac{V'(p)-V''(p)}{V'(p)}$$

the basic equation of conventional test theory. Notice, however, that the scaling constant, k, has been eliminated and the gain from testing is relative to the initial variance. This goes too far. We want to be able to compare the value of testing with the cost of testing and to be able to do this for many different situations. For these purposes, the expected value of counseling testing is

$$(23) \qquad EVCT = EU(\bar{p}',p)-EE_x U(\bar{p}''_x,p)$$
$$= k[V'(p)-V''(p)].$$

it should be understood that this equation for the expected gains resulting from testing for counseling assumes the use of an optimal estimation procedure. As will be shown below, not all estimation procedures used in counseling are optimal.

It is interesting to compute these values for the 10-item test described above. The first column in Table 15 gives the means for the conditional distribution of p for the various levels of guessing. These are, of course, the best possible estimates of an individual's ability level taking account both of the

Table 15.

Expected values for a 10-item test used for counseling decisions.

	x	$\bar{p}''$	$V(p)$	$E(\bar{p}''_0 - p)^2$	$\tilde{p}=\frac{x}{n}$	$\tilde{p}=\frac{1}{n}\left(\frac{x-n-x}{y}\right)$	$\tilde{p}=\frac{1}{n}(2x-n)$
	10	779	92		582	582	582
	9	723	107		421	339	166
	8	667	117		295	186	161
	7	611	125		204	127	570
	6	556	130		150	161	1394
$\theta = 0$	5	500	132		132	288	2632
	4	444	130		150	508	2105
	3	389	125		204	821	1638
	2	333	117		295	1228	1228
	1	277	107		421	876	876
	0	221	91		581	581	581
Expectation		500	124		227	457	408
	10	739	126	141	806	806	806
	9	675	144	166	648	542	299
	8	613	156	184	505	343	157
	7	553	162	196	379	215	400
	6	494	164	201	275	164	1030
$\theta = \frac{1}{5}$	5	439	160	198	198	201	2085
	4	386	152	185	153	338	1644
	3	338	139	165	154	593	1282
	2	294	124	140	213	990	990
	1	255	108	113	350	760	760
	0	221	89	89	577	577	577
Expectation		500	154	186	333	336	993
	10	658	201	345	1368	1368	1368
	9	586	212	398	1196	1045	668
	8	521	211	422	988	734	273
	7	463	202	421	762	464	242
	6	412	188	393	540	265	639
$\theta = \frac{1}{2}$	5	368	171	346	346	171	1524
	4	329	153	286	203	216	1237
	3	296	135	222	136	428	1012
	2	267	119	163	165	834	834
	1	242	101	114	303	687	687
	0	227	89	89	604	604	604
Expectation		500	200	394	875	672	624

distribution of ability levels in the population and of the guessing probability for the test. These estimates are graphed in Figure 6. Notice that a regression effect is apparent in these estimation procedures. For example, the highest possible test score does not imply that the person has the highest possible ability level while the lowest possible test score does not imply the lowest possible ability level. The effect is primarily due to the influence of the distribution of ability level in the population which, in this example, is symmetric about an average ability level. Therefore, if a person has an extreme test score, it is much more likely that his ability level is less extreme. This can be seen most clearly by examining the tables showing the conditional distributions of p given x contained in an earlier sub-section of this report.

Also graphed in Figure 6 are several other widely used estimates of a person's ability level. These estimates are either explicitly recommended or implied by many textbooks and test manuals. One estimate of an individual's ability level sometimes recommended and much more frequently used, is the proportion of test items passed, x/n. This is a straight line with slope of one graphed in Figure 6. The more sophisticated developers and users of tests have some appreciation of the effect of guessing and, thus, correct the test score, for chance before estimating an individual's ability level. They attempt to eliminate the effect of guessing by correcting the test score according to

$$\text{CORRECTED TEST SCORE} = R - \frac{W}{m-1}$$

where R is equal to the number of correct responses (equivalent to our x), W is the number of incorrect responses (equivalent to our n-x) and m is the number of possible answers listed in a multiple-choice item. Dividing this corrected test score by n, the total number of items in the test, yields an estimate of the person's ability level. Two such estimation schemes are graphed in Figure 6. One is for a five-alternative test which would have a minimum θ of 1/5; the other is for a two-alternative test with a miximum (and minimum) θ of 1/2.

Now let us consider the expected value of these various estimation procedures for the 10-item test affected by various levels of guessing as described previously. Figure 7 shows the expected value of counseling testing for a number of different estimates of a person's ability level. All of these estimates are derived from the person's test score, x, obtained by taking our 10-item test. Graph A shows the conditional expected values as a function of test score for each of the three

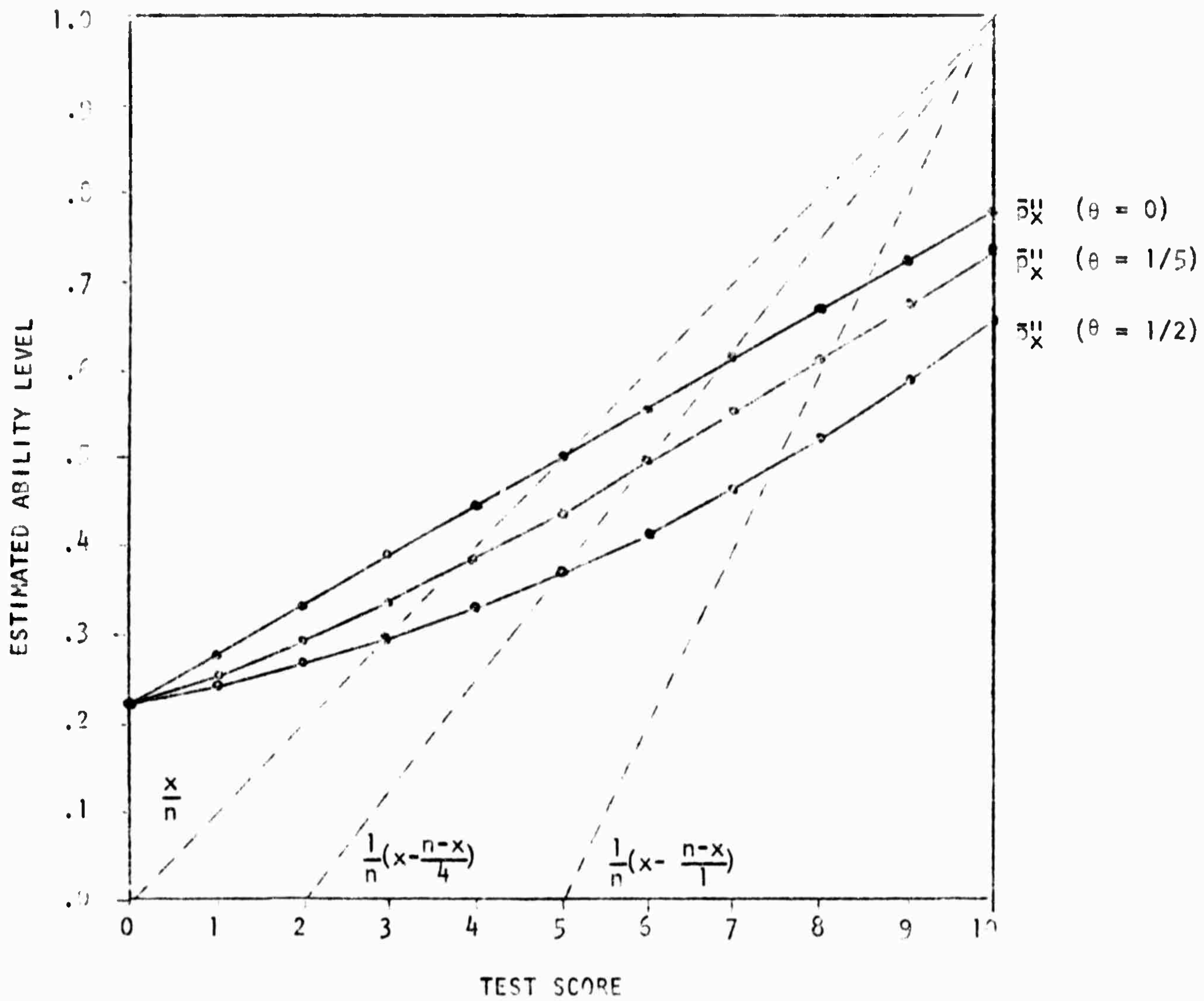


Figure 6. Various estimates of individual's ability level (p).

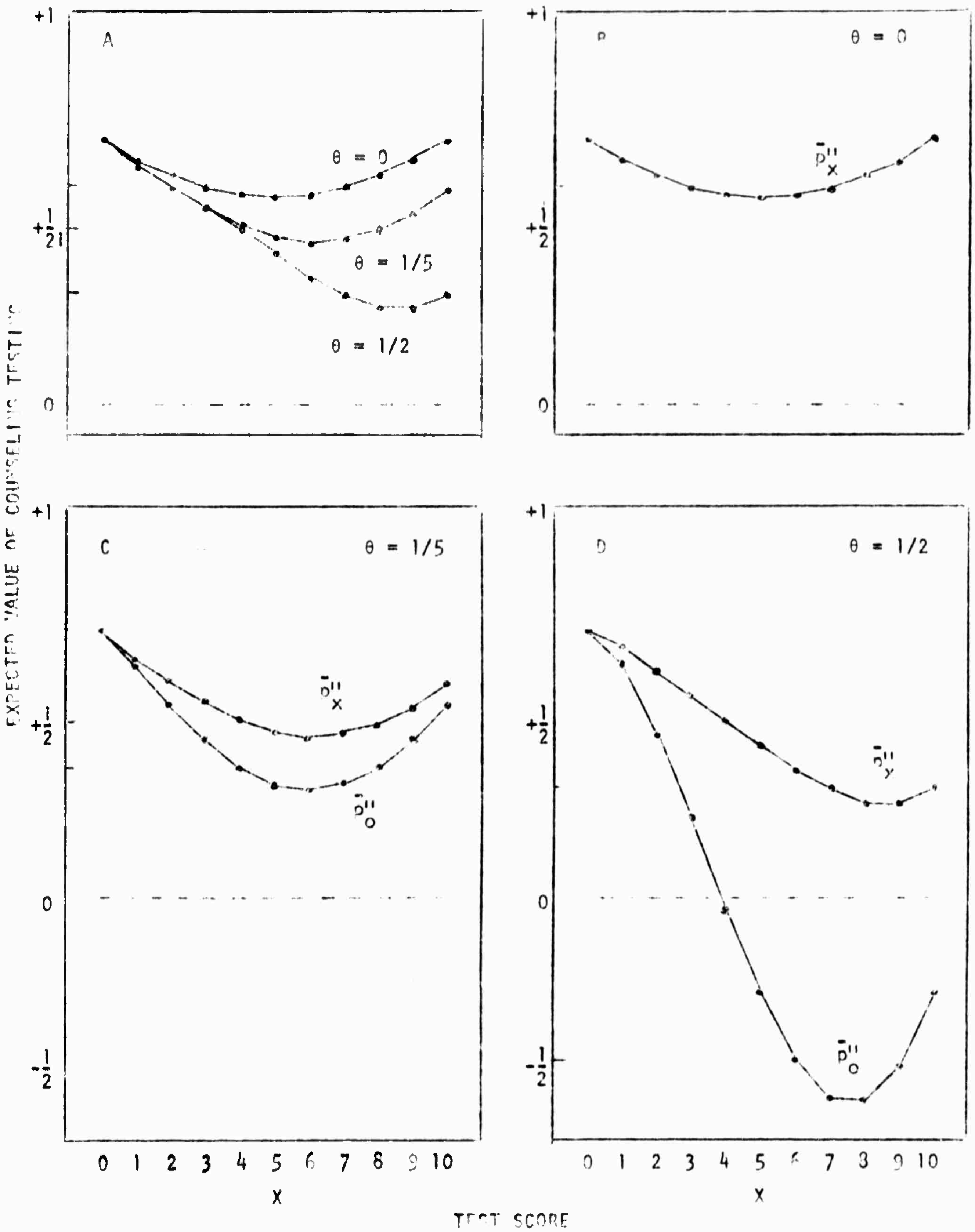


Figure 7. Expected values of counseling testing with a 10-item test affected by different degrees of guessing. Top of graph represents value of perfect information. Dashed line represents value of counseling without testing information.

optimal estimation procedures based upon the mean of the conditional distribution of p given x. The highest curve is obtained when there is no guessing in the test, the intermediate curve when there is a minimal level of guessing and the bottom curve when guessing is maximum. Notice that the expected values are depressed by the existence of guessing and that this depression is greater for the higher test scores. The dashed horizontal line corresponds to the expected value of counseling without testing and is set equal to zero on the scale of the ordinate. At the top of the graph, set equal to +1 is the value of perfect information which is obtained when the conditional variances are all equal to zero. Thus, as the length of the test is increased from the ten items to include all items in the very large pool of items, three curves would move toward the top of the graph approaching a horizontal line at +1. Also plotted in Graph A as pointers along the ordinate at the left are the overall expected values of counseling testing. The top marker, representing, of course, $\theta = 0$, the middle one, $\theta = 1/5$, and the bottom one, $\theta = 1/2$.

Graph B shows again the conditional expected values for the ten-item test with no guessing where the optimal estimate, β_x', is used to estimate a person's ability level. The top curve in Graph C is the same function, but for $\theta = 1/5$. If one ignored the existence of guessing and used as the estimate of an individual's ability level the mean of the conditional distribution of p given x based on the tables for $\theta = 0$, then the bottom curve shown in Graph C would be obtained. The use of this non-optimal estimation procedure would, of course, result in an additional loss in the expected value of counseling testing, but it is not too large in this case. Graph D, however, tells a different story. Here the top curve is for the optimal estimate, based upon the mean of the conditional distribution of p given x, when guessing is maximal, that is $\theta = 1/2$. The bottom curve shows the expected values of counseling testing using the non-optimal strategy which ignores the existence of guessing. Here the loss is great, so great, in fact, that it would be better to do no testing whatsoever and to estimate each person's ability level as being equal to the average ability level, β', of the population under consideration. In other words, the overall expected value of counseling testing by ignoring guessing in this case is negative implying that it is a poorer strategy than not testing at all.

Figure 8 shows the expected values of counseling testing for some other estimation procedures, for each of the three different levels of guessing. The top dashed curve in each case shows the conditional expected value of the optimal estimates as shown in Graph A of Figure 6. Notice that these have been plotted to a quite compressed scale. The dashed horizontal line at zero represents as

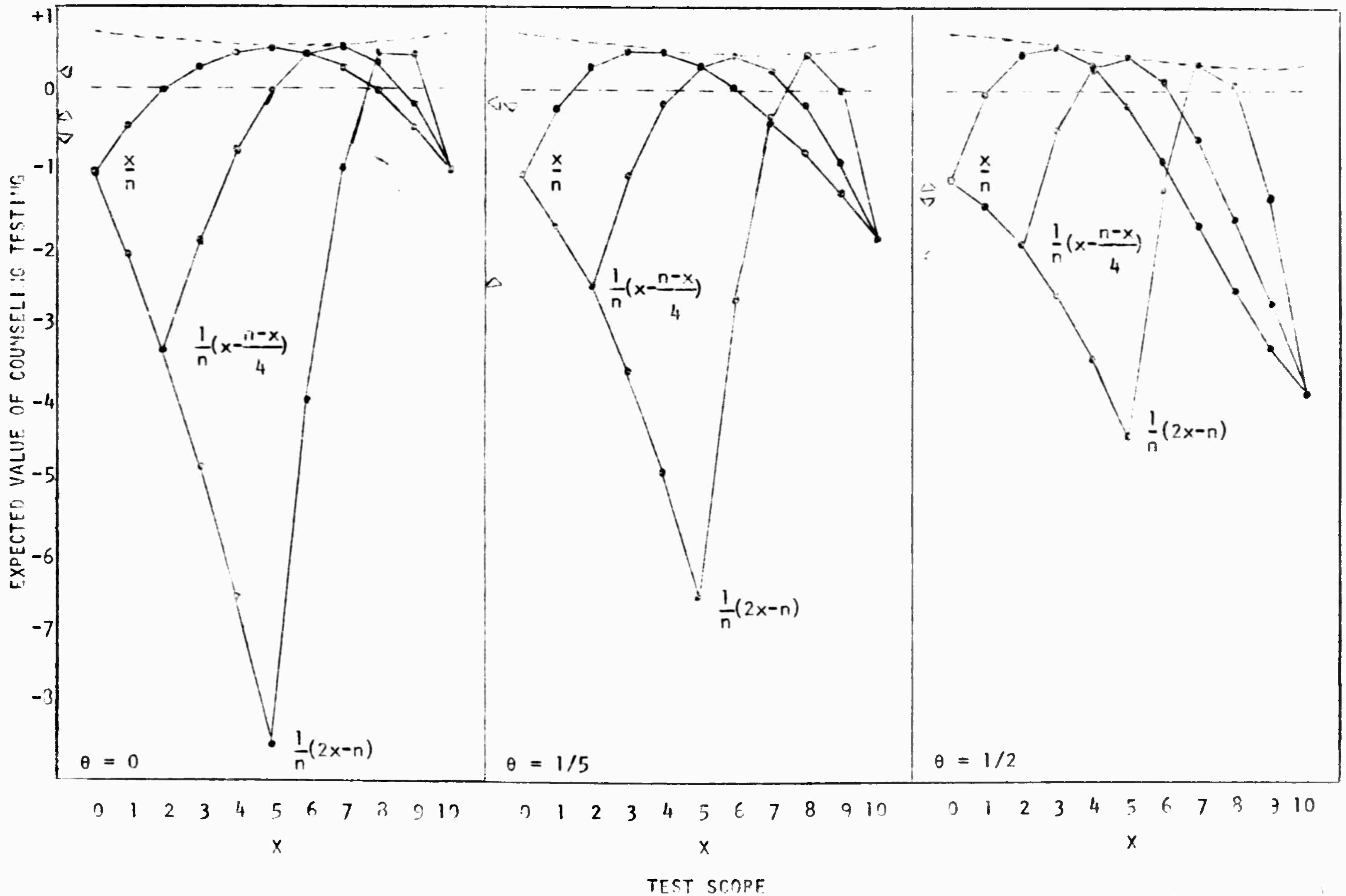


Figure 8. Expected values of counseling testing with a 10-item test affected by different degrees of guessing.

before the expected value of counseling without testing. In Figure 8 the three non-optimal strategies considered are (1) using the proportion of items passed as an estimate of a person's ability level, (2) correcting the test score for chance assuming a five-alternative test, and (3) correcting the test score for chance assuming a two-alternative test. The indices to the left near the ordinate represent the overall expected value of counseling testing with these non-optimal estimation procedures. With but one exception, they are all negative, implying that both the individual and the institution are much better served by not using testing information in counseling if this use is according to some of the frequently recommended procedures for estimating a person's ability level. The performance of these estimation procedures is really quite bad. And the way to see this independent of the absolute value of the utility scaling constant, k, is to realize that even giving a test of near infinite length which yields conditional variances of zero would just have the effect of moving the curves up nearer to the top of the graph. This is so because the plotted values are made up of two components, one being the conditional variance corresponding to uncertainty about the true mean, the other being the bias or the square of the difference between the conditional mean and the estimated ability level. Increasing the length of the test would reduce the size of the first component, but would not eliminate the second component. This can be seen by reexamining Figure 6. Increasing the length of the test makes the conditional mean, $\bar{p}''_x$, move closer to the diagonal corresponding to x/n, essentially eliminating the bias component between the conditional mean and the proportion of items passed. The two lines showing the correction for guessing, however, would not be affected and considerable bias would remain, even for a test of infinite length.

Again we find that the existence of guessing can have a serious effect upon the quality of personnel decisions. Ignoring the existence of guessing can lead to even further degradation in the value of counseling testing, sometimes of such magnitude that it is better not to use testing information in counseling. Ignoring the existence of information about the distribution of ability levels in the population leads to even greater degradation as does using the recommended formulas for correction for guessing. In most instances, the degradation is of such magnitude that that it is far better not to use testing information in counseling and to estimate each individual's ability level as being equal to the average ability level in the population.

RELIABILITY AND VALIDITY

Though correlational measures of test reliability and test validity are not very directly relevant to classification and counseling decisions, they are important to educational research and to behavioral science especially in those areas utilizing multivariate and factor-analytic techniques.

First we will define the maximum possible validity of the test as the correlation between test score and true ability level. In a sense this is the correlation between test score and true score in conventional test theory. It does not represent a test validity which is obtainable in practice, since it corresponds to the correlation between the test score and a perfectly measured criterion. It is, however, interesting to see what effect guessing has upon this maximum possible test validity. The computations are rather straightforward and are given in Table 16. The correlation is given in the column headed ρ

Table 16

Computations of correlation between test score and true ability level.

θ	$E(p)$	$E(x)$	$E(p^2)$	$E(x^2)$	$E(xp)$	$V(p)=E(p^2)-[E(p)]^2$	$V(x)=E(x^2)-[E(x)]^2$	$Cov(xp)=E(xp)-E(x)E(p)$	$\rho=\frac{Cov(xp)}{\sqrt{V(x)V(p)}}$	ρ^2
0	.5	5.0	.27782	30.00	2.77807	.027824	5.0	.27807	.745	.555
1/5	.5	6.0	.27782	40.00	3.22255	.027824	4.0	.22242	.667	.445
1/2	.5	7.5	.27782	58.75	3.89239	.027824	2.5	.14239	.540	.292

and is degraded by guessing from an initial value of about .7 down to a value maximally affected by guessing of about .5. ρ^2 is used in conventional test theory to measure the percentage of variance accounted for by the test. These values of ρ^2 agree closely with the percentage reduction in variance computed from the expected variances of Table 15 and range from about .5 down to about .3.

Second, we define test reliability as the correlation between the scores obtained from two tests, each test being made up of a set of items randomly sampled from the pool of test items. This measure corresponds to the correlation between equivalent test forms in conventional test theory. It is instructive to compute the test reliability for our ten-item test described above. Realize that

Table 17.

Joint distributions of x and y. Entries to be scaled times 10^{-3}.

	1st Score, x	2nd Score, y: 0	1	2	3	4	5	6	7	8	9	10
No guessing ($\theta = 0$)	10						1	2	3	4	4	2
	9				1	2	4	7	9	10	8	4
	8			1	3	6	11	16	18	16	10	4
	7		1	3	7	14	20	25	24	18	9	3
	6		2	6	14	22	28	30	25	16	7	2
	5	1	4	11	20	28	32	28	20	11	4	1
	4	2	7	16	25	30	28	22	14	6	2	
	3	3	9	18	24	25	20	14	7	3	1	
	2	4	10	16	18	16	11	6	3	1		
	1	4	8	10	9	7	4	2	1			
	0	2	4	4	3	2	1					
Minimum guessing ($\theta = 1/5$)	10					1	1	3	5	7	7	4
	9				1	3	6	12	17	19	15	7
	8			1	3	8	16	24	30	28	19	7
	7		1	3	7	16	26	35	37	30	17	5
	6		1	5	12	22	33	38	35	24	12	3
	5		2	6	15	25	32	33	26	16	6	1
	4		2	7	14	21	25	22	16	8	3	1
	3		2	6	11	14	15	12	7	3	1	
	2		1	3	6	7	6	5	3	1		
	1		1	1	2	2	2	1	1			
	0											
Maximum guessing ($\theta = 1/5$)	10					1	2	6	14	23	27	19
	9				1	3	9	20	37	52	51	27
	8				1	6	15	32	52	63	52	23
	7				2	7	18	34	49	52	37	14
	6				2	6	14	26	34	32	20	6
	5				1	4	9	14	18	15	9	2
	4				1	2	4	6	7	6	3	1
	3					1	1	2	2	1	1	
	2											
	1											
	0											

because of the independent random sampling of the two sets of ten items, the test items themselves are independent. This independence holds, however, only at each fixed ability level. To see this, consider computing the joint distribution of test scores, x, from the first test and test scores, y, from the second test. For, given p, the joint probability of x and y is given by

$$P(x,y|p) = P(x|p)P(y|p). \quad (24)$$

Thus, for each value of p, we have a table of joint probabilities with a zero correlation between x and y. When we sum over p, however, to obtain an unconditional joint probability of x and y,

$$P(x,y) = \sum_p P(x|p)P(y|p)P(p) \quad (25)$$

we end up with a table of joint probabilities for x and y which are positively correlated. These joint distributions of x and y are given in Table 17 for the different levels of guessing. Notice that the effect of increasing amounts of guessing is to concentrate the distribution in the positive quadrant.

These joint probabilities (but to more decimal places) have been used to compute the correlation between x and y and are shown in Table 18. Here the test

Table 18

Computations of correlation between test scores from equivalent forms.

θ	$E(y)$	$E(x)$	$E(y^2)$	$E(x^2)$	$E(xy)$	$V(y)=E(y^2)-[E(y)]^2$	$V(x)=E(x^2)-[E(x)]^2$	$Cov(xy)=E(xy)-E(x)E(y)$	$\rho=\frac{Cov(xy)}{\sqrt{V(x)V(y)}}$	ρ^2
0	5.0	5.0	30.00	30.00	27.78239	5.0	5.0	2.78245	.556	.309
1/5	6.0	6.0	40.00	40.00	37.78095	4.0	4.0	1.77999	.445	.198
1/2	7.5	7.5	58.75	58.75	57.00397	2.5	2.5	.50689	.203	.041

reliability ranges from about .56 down to .20 for maximum guessing while the percentage of variance accounted for by the test ranges from about .31 down to about .04. This seems to be a fairly significant degradation in test reliability due to the effect of guessing.

These test reliabilities are a function both of the length of the test and, more subtly, the distribution of ability levels in the population, since the variance of ability level moderates the correlation. A future report will give test reliabilities for different test lengths and distributions of ability level. Though the test reliabilities will undoubtedly increase with increases in test length, the degradation due to the guessing should not be entirely discounted. As Cronbach and Gleser mention, in their discussion of the band-width fidelity paradox, increasing the band-width of a test can greatly improve the quality of personnel decisions. Thus, personnel testing should move in the direction of using test batteries made of many short tests, each one presumably measuring a different dimension. Thus, it would seem that the gain in reliability, due to the elimination of guessing would be of great importance to the success of these wide band-width procedures.

TESTWISENESS

And finally we come to a consideration of individual test-taking strategies. Thus far we have been concerned with the effect of guessing upon the quality of institutional decisions. It is also possible to analyze the effect of guessing from the point of view of individual decisions. For example, in many situations in which an individual takes a test, he would like very much to make a high score. He wants this high score because he knows that it is necessary if he is to be admitted to college by passing a college entrance exam, to be allowed to begin a career as a result of qualifying on a federal or state civil service examination, or to be enrolled in a special manpower development or Job Corps training program. Now in these and many other instances of testing it is quite important to the individual to achieve a test score exceeding the cutting score. If he does, he will get a chance to achieve his goals; if he does not he is denied any opportunity to do so. In these types of situations, the utility to the individual of the various courses of action can be measured effectively by the probability of his test score exceeding a cut-off score. This is the probability of his achieving a desired goal and the expected utility of a particular course of action is proportional to this probability. Therefore, it is interesting to compute this probability for our ten-item test described previously.

Given that an individual has an ability level, p, his probability of passing a test with cutting score, c, is

(26) $$P(x \geq c|p,n) = \sum_{x=c}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

This is just the cumulative of the binomial distribution derived above and is the appropriate equation to be used when the test is unaffected by guessing. However, when guessing can occur on the test, as, for example, when a conventional choice test is used, the following equation is appropriate

(27) $$P(x \geq c|(1-\theta)p+\theta,n) = \sum_{x=c}^{n} \binom{n}{x} [(1-\theta)p+\theta]^x[1-\{(1-\theta)p+\theta\}]^{n-x}$$

where θ is the individual's probability of getting correct through guessing those items which he does not know. Remember that according to Theorem I guessing just has the effect of increasing the probability parameter in the binomial distribution.

Now, one decision that must be made by an individual taking a test is whether or not to guess at the answer to those items that he does not know or, more generally, on which of the items should he guess. Let us just consider the extremes of these test-taking strategies, that is, the individual decides between guessing at those items that he does not know or not guessing at any of these items. Thus, if an individual of ability level, p, chooses to guess on the test and his probability of guessing the correct answer is θ, then his chance of passing the test is given by Equation 27. If, on the other hand, the individual chooses not to guess at those items that he does not know, his chance of passing the test is given by Equation 26. The value of Equation 26 will always be less than the value of Equation, 27, since the probability parameter of the binomial distribution is smaller in the first case. Thus, if we subtract Equation 26 from Equation 27 we obtain the measure of how much an individual's chances of passing a test have been reduced by his refusing to guess. Under the conditions described above, the individual's expected utility is reduced proportionately. Figure 9 shows the reduction in chance of passing the test as a result of not guessing for different ability levels and for different cutting scores. For a cutting score of 9 and a θ of 1/5, the reduction is moderate, achieving its maximum of .13 at an ability level of .8. When the probability of getting an item correct by chance increases to 1/2, however, the losses become more significant and achieve a maximum of about .40 at an ability level of .7. The loss is much less among the lower ability levels. This is primarily due to the fact that these individuals have essentially no chance of passing the test no matter what they do. Observe that a ten-item test with a cutting score of 9 would be designed to select out only the most

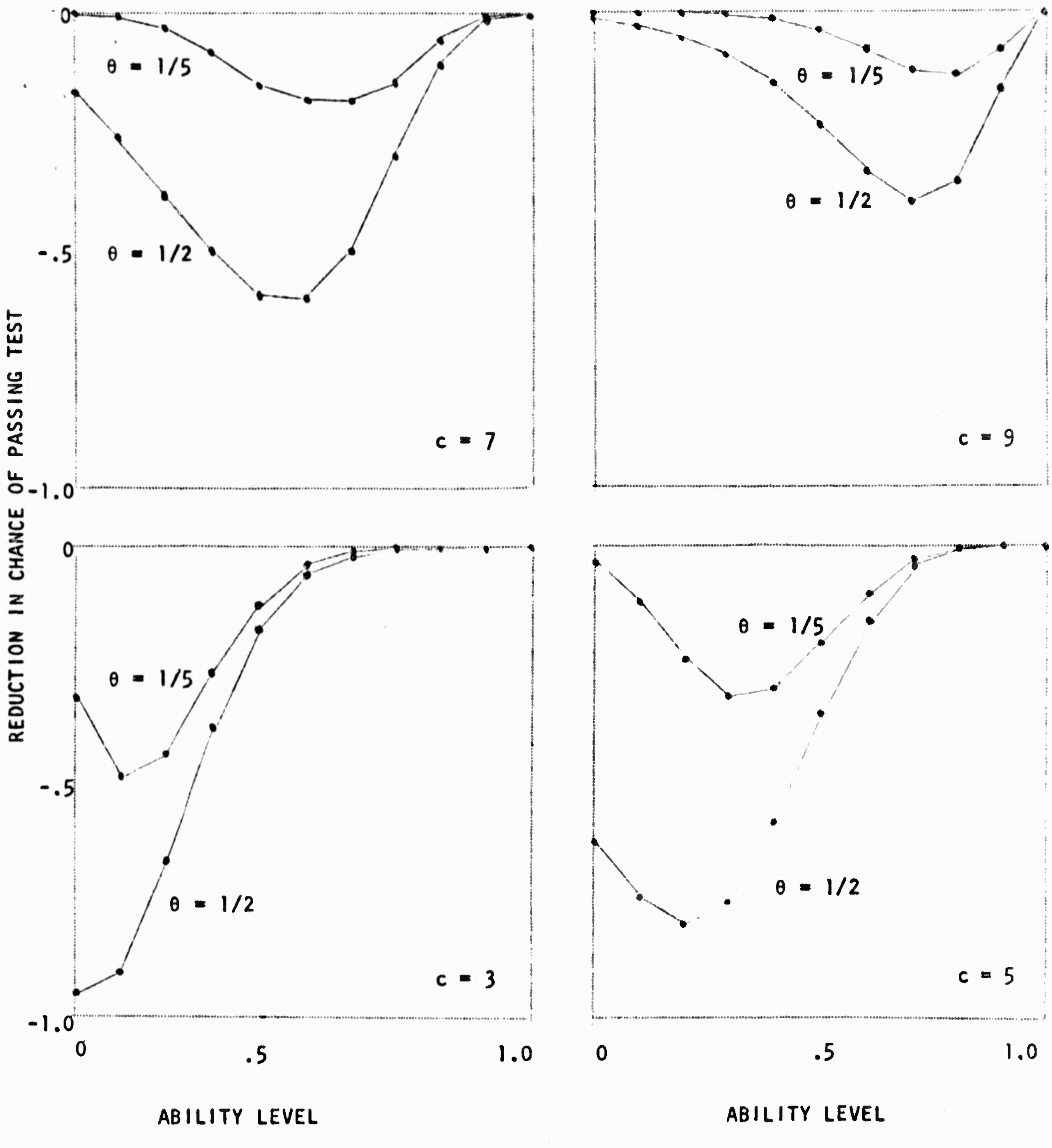


Figure 9. Absolute reduction in chance of passing tests affected by two degrees of guessing.

competent individuals in the tested population. Observe also that the effect of guessing tends to be large among just these individuals.

A cutting score of 7 on a ten-item test is more or less comparable to passing at the 70% level in some educational tests. Here the loss due to not guessing becomes more significant with the maximum for $\theta = 1/5$ being about .21 at an ability level of .6 while the maximum for $\theta = 1/2$ is about .60 at an ability level of .5. Here again, the effect of guessing or not guessing appears to be large for those individuals who are borderline with respect to the cut-off point of the test.

A cutting score of .5 on a ten-item test may serve to illustrate the use of an educational test to divide individuals into two groups for further instruction more closely tailored to their ability levels. Here the effect of guessing is becoming even larger with the maximum loss for $\theta = 1/5$ being about .32 for an ability level of .3 and in the case of $\theta = 1/2$, the maximum loss is about .80 for an ability level of .2. In this case, certainly, whether or not an individual guesses on the test can become the major factor determining in which of the two instructional programs he is placed.

Finally, we have a ten-item test with a cutting score of 3 which might represent the use of a test to screen out only the lowest ability individuals and prevent them from entering some program. Here the effect of guessing or not guessing is greater yet with a maximum for $\theta = 1/5$ being approximately .49 for an ability level of .1 and in the case of $\theta = 1/2$ the maximum loss is about .95 at an ability level of 0. In testing for this type of purpose, it would seem that if some of the individuals were guessing and others were not guessing, then who would be screened out of the program would be largely determined by whether or not the individual chose to guess in taking the test.

In review, these differences are quite large. With such large differences, it would not be surprising that individuals would learn as a result of taking conventional choice tests that their chances are much improved by always guessing at those items which they do not know. With sufficient experience taking conventional choice tests, they might even learn to ignore instructions to the effect thay they should not guess and proceed to guess anyway. Remember that even in the case of scoring systems designed to penalize guessing we have found that they do not actually work and that the individual suffers no loss from guessing. Therefore, let us define testwiseness this way. The testwise individual will always guess at the answers to those items which he does not know. Other, less experienced individuals may obey test instructions to the effect that they should not guess or for some other reason such as an aversion to gambling or "faking", they will not guess at the answers to those items that they do not know. It is interesting to conjecture that the proportion of non-testwise individuals who refure to guess is much larger among the dropouts and the educationally disadvantaged. This, taken together with the results in this report makes one wonder whether the use of conventional choice testing with these individuals may not yield very biased information and unfair decisions with respect to their futures.

REFERENCES

Coombs, C. H., Milholland, J. E. & Womer, F. B. (1955) The assessment of partial knowledge. J. ed. psychol. Meas., 16, 13-37.

Cronbach, L. J. & Gleser, G. C. (1965) Psychological tests and personnel decisions. Urbana: University of Illinois Press.

Ebel, R. L. (1965) Measuring educational achievement. Englewood Cliffs: Prentice-Hall.

de Finetti, B. (1937) La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré, 7. [Translated and reprinted as "Foresight: its logical laws, its subjective sources" in H. E. Kyburg, Jr. & H. E. Smokler (Eds.) Studies in subjective probabilities. New York: Wiley, 1964]

de Finetti, B. (1962) Does it make sense to speak of good probability appraisers? In I. J. Good (Gen. Ed.) The scientist speculates. New York: Basic Books, 357-364.

Massengill, H. E. & Shuford, E. H., Jr. (1965) Direct vs. indirect assessment of simple knowledge structures. ESD-TR-65-542, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

van Naerssen, R. F. (1961) A scale for the measurement of subjective probability. Acta Psychologica, 159-166.

Raiffa, H. & Schlaifer, R. (1961) Applied statistical decision theory. Boston: Division of Research, Harvard Business School.

Ramsey, F. P. (1926) The foundation of mathematics and other logical essays. New York: Humanities Press.

Roby, T. B. (1965) Belief states: a preliminary empirical study. ESD-TDR-64-238, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Savage, L. J. (1954) The foundations of statistics. New York: Wiley.

Shuford, E. H., Jr. (1964) Some Bayesian learning processes. In Shelly, M. W. & Bryan, G. L. (Eds.) Human judgments and optimality. New York: Wiley, pp. 127-152.

Shuford, E. H., Jr. (1965) Cybernetic testing. ESD-TR-65-467, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Shuford, E. H., Jr., Albert, A., & Massengill, H. E. (1966) Admissible probability measurement procedures. Psychometrika, 31, 125-145.

Shuford, E. H., Jr. & Massengill, H. E. (1965) On communication and control in the educational process. ESD-TR-65-568, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Toda, M. (1963) Measurement of subjective probability distributions. ESD-TDR-63-407, Decision Sciences Laboratory, L. G. Hanscom Field, Bedford, Mass.

Toda, M. & Shuford, E. H., Jr. (1965) Logic of systems: introduction to the formal theory of structure. In General systems theory yearbook: 1965.

Toda, M. & Shuford, E. H., Jr. (1965) Utility, induced utilites, and small worlds. Behavioral Science, 10, 238-254.

Watanab , S. M. (1960) Information theoretical aspects of inductive and deductive inference. IBM J. Research and Develop., 4.

Watanabe, S. M. (1966) A quantative study of certain formal aspects of knowing and guessing. New York: Wiley, (In press).

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) The Shuford-Massengill Corporation Box 26 Lexington, Massachusetts 02173	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

DECISION-THEORETIC PSYCHOMETRICS: THE EFFECT OF GUESSING ON THE QUALITY OF PERSONNEL AND COUNSELING DECISIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)
Emir H. Shuford, Jr.
H. Edward Massengill

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November 1966	53	20

8a. CONTRACT OR GRANT NO AF 49(638)1744	9a. ORIGINATOR'S REPORT NUMBER(S) SMC R-5
b. PROJECT NO 920F-9719	
c. 6154501R	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR-68-2155
d. 681313	

10. DISTRIBUTION STATEMENT
1. This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
TECH, OTHER	Air Force Office of Scientific Research 1400 Wilson Boulevard (SRLB) Arlington, Virginia 22209

13. ABSTRACT

In Section B, The Effect of Guessing on the Quality of Personnel and Counseling Decisions, the fundamental probability distributions for total test scores are derived by assuming that each person knows the answers to some items and guesses on the remaining items. Analysis of a 10-item test shows that guessing levels encountered in practice (a) seriously degrade the value of selection, placement, and counseling decisions, (b) significantly impair test reliability and validity, and (c) magnify the influence of testwiseness.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Guessing						
Test-taking strategies						
Counseling Decisions						