中華心理學刊
民86，39卷，2期，173-192

Chinese Journal of Psychology
1997, Vol. 39, No. 2, 173-192

SOME NEW RESULTS ON HIT RATES AND BASE RATES IN MENTAL TESTING

PETER H. SCHÖNEMANN
Department of Psychology
*National Taiwan University**

Recent work on hit rates and base rates (Schönemann and Thompson, 1996) is extended: A flawed premise in the derivation of an earlier hit rate approximation, HR1, is corrected, leading to a slightly more complicated approximation, HR2. However, over the targeted parameter region, the differences between HR1 and HR2 are small.

After deriving exact hit rates for 2×2 contingency tables with binary criteria, they are compared with HR1 and HR2, and also with hit rates for continuous criteria inferred, via Bayes' Theorem, from Taylor and Russell's (1939) tables. Overall, the simpler approximation HR1 outperforms HR2.

Finally, a new approximation is derived for the minimum validity needed that a test improves over random admissions in terms of total percent of correct classifications. More than four decades ago, Meehl and Rosen (1955) warned that validity coefficients, in isolation, are insufficient for gauging the practical merit of a test, because, "... when the base rates of the criterion classification deviate greatly from a 50 percent split, use of a test sign having slight or moderate validity will result in an increase of erroneous clinical decisions." (p. 215. Emphasis in the original). The present results corroborate these concerns.

Keywords: Mental tests, IQ tests, Hit rate bias, Base rate problems, Predictive validities

1. Introduction

This paper extends previous work on hit rates and base rates by Schönemann and Thompson (1996). Our interest in these problems had been aroused when we reanalyzed data sets the NCAA (Note 1) had collected in support of a projected upward revision of standards to qualify for athletic scholarships.

Inspection of these data revealed a disproportionate number of "False negatives" for Blacks, compared to Whites. "False negatives" are false classifications of qualified candidates as "unqualified", because they do not pass the admission test. For fallible tests, some such classification errors are virtually unavoidable, as are the complementary misclassifications of unqualified candidates as "qualified", because they pass the admissions test ("false positives"). However, what drew our attention was that the error

---

On leave from Purdue University. The bulk of this work was done during my visits to Taiwan in Summer 1996 and Spring 1997, which were supported by the Foundation for the Advancement of Outstanding Scholarship (FAOS). I wish to thank Professors Sigmund Hsiao, Huang Jong-Tsun, Liu In-Mao, and Jeanie Weng for their active support on my behalf and all my colleagues at Taida for their warm hospitality.

Please address reprint request and correspondence to Peter H. Schönemann, Department of Psychological Sciences, Purdue University, West Lafayette IN 47907, USA. Email: phs@psych.purdue.edu.

rates, more specifically the

"Miss Rate":= Proportion of qualified students failing the test

and also the

"False Alarm Rate":= Proportion of unqualified students passing the test

differed systematically between Whites and Blacks: Black miss rates consistently exceeded White miss rates, and White false alarm rates consistently exceeded Black false alarm rates. Thus, unqualified Whites benefited from the test errors at the expense of qualified Blacks. Subsequent analyses described in more detail in (Schönemann and Thompson, 1996) showed that similar asymmetries characterize rich/poor comparisons. In an early paper, Cole (1973) has argued that this constitutes a form of bias against Blacks (in this case), who face steeper odds than Whites to acquire a decent education to begin with, and then face an additional hurdle at the college entrance stage because the admission tests systematically screen out higher proportions of qualified Blacks than Whites, and screen in a higher proportions of unqualified Whites than Blacks. As Hartigan and Wigdor (1988) put it: "Fair test use would seem to require at the very least that the inadequacies of the technology should not fall more heavily on the social groups already burdened by the effects of past and present discrimination" (p. 260).

2. Terminology and Notation

To render the discussion of these issues manageable, the notation laid out in Table 1 will be used throughout this paper.

On the left side of Table 1, a 2×2 joint probability table is laid out. The columns represent the actual qualifications of the applicants: A candidate is either qualified (Q) or unqualified (U). For entrance tests, "qualified" might mean graduation with a BA, "unqualified" failure to receive a BA. Criteria such as these with only two outcomes are called "binary", in contrast to "continuous" criteria such as Freshman gradepoint average (FGPA).

To predict qualification at the ad-mission stage, a test is given. In Table 1 the two test outcomes are represented by the rows of the 2×2 joint probability table: A candidate either passes (P) or fails (F) the admission test. As a result, the table contains four joint probabilities, two for correct decisions and two for false decisions. The correct decisions are called "true positives" (with joint probability tp) and "true negatives" (tn), the incorrect decisions, "false positives" (fp, unqualified candidates who pass the test) and "false negatives" (fn, qualified candidates who fail the test). All four cells sum to 1.

The column sum tp+fn gives the proportion of qualified candidates in the unselected population. It is called a "base rate" (b). Similarly, the column sum fp+tn (=1−b) is the

Table 1
Definitions and notation

A. Probability Tables:

	(a) Joint Criterion:			(b) Conditional Criterion:	
	U	Q		U	Q
Test:					
P	fp	tp	q	fr	hr
Cut-off (C)					
F	tn	fn	1−q	1−fr	1−hr
	1−b	b	1.	1.	1.

Outcomes:
Criterion: U:=unqualified Q:=qualified (e.g., graduates)
Test: P:=passes test F:=fails test

B. Notation for Joint Probabilities:
fp:=false positive tp:=true positive tn:=true negative fn:=false negative
q:=(admission) quota b:=base rate %c:=tp+tn=Proportion of Correct Decisions
fp+tp+tn+fn=1
q:=fp+tp b:=tp+fn

C. Notation for Conditional Probabilities:
fr:=fp/(1−b)=false alarm rate hr:=tp/b=hit rate
sr:= tp/q=success rate (Taylor-Russell probabilities)

complementary base rate, the proportion of unqualified subjects in the population. The row sum fp+tp is the proportion of candidates who pass the test, and thus, a function of the cut-off C defining "pass" on the test. It will be called the (admission-) "quota"(q).

The quota is a characteristic of the test, and, thus, under the control of the tester, who can raise or lower q by adjusting the test cut-off C. The base rate is a property of the population and thus not under his control. One lesson to emerge - already stressed by Meehl and Rosen (1955) but conveniently ignored ever since - is that the practical merit of a test is not just a function of its predictive validity (test-criterion correlation r), but also of the base rate b and the quota q.

On dividing out the base rates (i.e., on conditioning the joint table on its columns), one arrives at the table of conditional probabilities on the right in Table 1. Of particular interest here will be the

(1) "hit rate": hr:=tp/b.

As already noted, it represents the proportion of qualified candidates who pass the test. Ideally, hr should be close to 1, but for fallible tests it may fall far short of this ideal. On forming the ratio of hit rates for Whites versus hit rates for Blacks, HRB:= hr_W/hr_B, one obtains a measure of hit rate bias against Blacks. On reanalyzing several data sets, we found that, for college entrance tests, this bias averages out to 1.7. This means, roughly, that *qualified* Whites are twice as likely to pass the test as *qualified* Blacks.

3. Earlier Theoretical Results

In (Schönemann and Thompson, 1996) we derived, as a simple consequence of the definitions in Table 1, a Hit Rate Bound:

(2) $hr \leq q/b$,

which means that simply raising a test cut-off ("raising standards") does not ensure an improvement in correct decisions: Especially if the base rate is large, it may just raise the proportion of misses (since mr=1−hr). Thus, depending on one's objective, one may wish to exercise control over hit rates. This is made difficult by the fact that the relation between hr, b, q, and r cannot be solved explictly for hr as a function of the other three parameters. However, we were able to show that for binary criteria a simple and, over the relevant parameter region quite reasonable, hit rate approximation is given by

(3) $HR1 := q + r_{pb}\sqrt{(1-b)/b'}/3$,

where "r_{pb}" denotes the point biserial correlation measuring the validity of the test.

We derived this explicit hit rate esti-mate HR1 by approximating the standard normal ogive by a straight line with slope 1/3 over the base rate interval ($.3 \leq b \leq .7$). Our rationale for this restriction was that, beyond this range, the misclassification rates quickly reach unacceptable levels for all tests except those with purely academic validities (e.g., Estes, 1992. Note 2). More generally, all approximations considered here are intended only for the practically relevant parameter region

(4) $(0 \leq r \leq .5)$, $(.3 \leq b \leq .7)$, $(.3 \leq q \leq .7)$.

4. Derivation of Exact Hit Rates

Our derivation of HR1 are briefly reviewed in Appendix 1. As noted there, in loc. cit. we used the within group standard deviation (s_w) to norm the mean distance between the qualified and the unqualified group in strict analogy with Signal Detection Theory (SDT). However, in the present context, the total standard deviation

(s_t) is more appropriate. As shown in Appendix 1, this revision yields an "improved" estimate

(5) $$HR2{:=}q+(r/\sqrt{1-r^2})\sqrt{(1-b)/b}\,/3,$$

which differs from HR1 in the factor $k{:=}\ r/\sqrt{1-r^2}$. For small r, the effect of k is negligible, so that one expects that the "improved" hit rate approximation HR2 will be close to HR1.

To obtain "exact" hit rates (Note 3), the above correction was applied to the point biserial formula,

(6) $$r_{pb}{=}d''\sqrt{b(1-b)},$$

where $d''{:=}(m_Q{-}m_U)/s_t$ is now the correctly normed mean difference. As shown in Appendix 1, this leads to

(7) $$r_{pb}{=}\sqrt{r_{pb}{'}^2/(1+r_{pb}{'}^2)},$$

where $r_{pb}'{:=}d'\sqrt{b(1-b)}$ is the point biserial based on norming with s_w. Again it is clear that the revision takes effect only for r_{pb}'s near the upper boundary .5 (Note 4).

After programming (7), "exact" hr's were computed iteratively. For fixed b and q, hr was varied until a specified r_{pb} was reproduced within a small tolerance (.002). The results are tabulated in Tables 2 and 3.

These "exact" hit rates were then compared with (a) the simpler HR1 (Table 2), (b) the "improved" approximation HR2 (Table 3), and (c), also, the "exact" hit rates inferred via Bayes' Theorem from Taylor and Russell's (1939) tables of the "success rates" (sr := tp/q. Note 5) for continuous criteria, viz: pr(pass|qualified) = pr(qualified|pass) pr(pass)/pr(qualified)

(8) $$hr{=}sr{\times}q/b.$$

The results of these comparisons are summarized in Tables 4 and 5.

As the columns of differences in these tables show, the simpler approximation HR1 - though strictly speaking derived from a flawed premise - outperforms the "improved" approximation HR2 both for binary and for continuous criteria over the targeted parameter ranges (eq. 4). For binary criteria, the largest discrepancy is .07, with the modal discrepancy near .03. Since only 3 out of 125 discrepancies over the targeted parameter range are negative, the approximation could be further improved by raising the multiplier (1/3) slightly. This seems hardly worthwhile since the "improved" HR2 is only slightly better than HR1 for binary criteria.

Moreover, for continuous criteria, HR1 overestimates the "exact hr's inferred from

Table 2
"Exact" hit rates for binary criteria, HR1, and differences

		Exact						HR1						Differences					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	25	29	34	40	46	53	25	30	35	40	46	51	0	-1	-1	0	1	2
30	30	35	41	47	54	61	70	35	40	45	50	56	61	0	1	2	4	6	9
30	40	46	52	58	66	73	81	45	50	55	60	66	71	1	2	3	6	8	9
30	50	56	62	69	75	82	89	55	60	65	70	76	81	1	2	4	5	7	8
30	60	66	72	77	83	89	94	65	70	75	80	86	91	1	2	2	3	4	3
30	70	75	80	85	89	94	97	75	80	85	90	96	101	0	0	0	-1	-2	-4
30	80	84	88	91	94	97	99	85	90	95	100	105	111	-1	-2	-4	-6	-8	-12
40	20	24	27	31	35	39	43	24	28	32	36	40	45	0	-1	-1	-1	-1	-2
40	30	34	39	44	49	54	60	34	38	42	46	50	55	0	1	2	3	4	6
40	40	45	50	55	60	66	73	44	48	52	56	60	65	1	2	3	4	6	9
40	50	55	60	65	71	77	83	54	58	62	66	70	75	1	2	3	5	7	9
40	60	65	70	74	79	85	90	64	68	72	76	80	85	1	2	2	3	5	6
40	70	74	78	83	87	91	95	74	78	82	86	90	95	0	0	1	1	1	1
40	80	83	87	90	93	96	98	84	88	92	96	100	104	-1	-1	-2	-3	-4	-6
50	20	23	26	29	31	34	37	23	27	30	33	37	40	0	-1	-1	-2	-3	-3
50	30	34	37	41	44	48	52	33	37	40	43	47	50	1	0	1	1	1	2
50	40	44	48	52	56	61	66	43	47	50	53	57	60	1	1	2	3	4	6
50	50	54	58	62	67	72	77	53	57	60	63	67	70	1	1	2	4	5	7
50	60	64	68	72	76	81	86	63	67	70	73	77	80	1	1	2	3	4	6
50	70	74	77	81	84	88	93	73	77	80	83	87	90	1	0	1	1	1	3
50	80	83	86	89	91	94	97	83	87	90	93	97	100	0	-1	-1	-2	-3	-3
60	20	22	24	26	29	30	32	23	25	28	31	34	36	-1	-1	-2	-2	-4	-4
60	30	33	36	38	41	44	47	33	35	38	41	44	46	0	1	0	0	0	1
60	40	43	46	50	53	56	60	43	45	48	51	54	56	0	1	2	2	2	4
60	50	53	57	60	64	68	72	53	55	58	61	64	66	0	2	2	3	4	6
60	60	63	67	70	74	78	82	63	65	68	71	74	76	0	2	2	3	4	6
60	70	73	76	79	82	86	90	73	75	78	81	84	86	0	1	1	1 2	4	6
60	80	82	85	87	90	93	96	83	85	88	91	94	96	-1	0	-1	-1	-1	0
70	20	22	23	25	26	27	28	22	24	26	29	31	33	0	-1	-1	-3	-4	-5
70	30	32	34	36	38	40	42	32	34	37	39	41	43	0	0	-1	-1	-1	-1
70	40	43	45	47	50	52	54	42	44	47	49	51	53	1	1	1	1	1	1
70	50	53	55	58	61	64	67	52	54	57	59	61	63	1	1	2	2	3	4
70	60	63	65	68	71	74	78	62	64	67	69	71	73	1	1	2	2	3	5
70	70	72	75	77	80	83	87	72	74	77	79	81	83	0	1	1	1	2	4
70	80	82	84	86	89	91	94	82	84	87	89	91	93	0	0	-1	0	0	1

Decimal points omitted.

Table 3
"Exact" hit rates for binary criteria, HR2, and differences

		Exact						HR2						Differences					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	25	29	34	40	46	53	25	30	36	42	49	58	0	-1	-2	-2	-3	-5
30	30	35	41	47	54	61	70	35	40	46	52	59	68	0	1	1	2	2	2
30	40	46	52	58	66	73	81	45	50	56	62	69	78	1	2	2	4	4	3
30	50	56	62	69	75	82	89	55	60	66	72	79	88	1	2	3	3	3	1
30	60	66	72	77	83	89	94	65	70	76	82	89	98	1	2	1	1	0	-4
30	70	75	80	85	89	94	97	75	80	86	92	99	108	0	0	-1	-3	-5	-11
30	80	84	88	91	94	97	99	85	90	96	102	109	118	-1	-2	-5	-8	-12	-19
40	20	24	27	31	35	39	43	24	28	33	38	44	51	0	-1	-2	-3	-5	-8
40	30	34	39	44	49	54	60	34	38	43	48	54	61	0	1	1	1	0	-1
40	40	45	50	55	60	66	73	44	48	53	58	64	71	1	2	2	2	2	2
40	50	55	60	65	71	77	83	54	58	63	68	74	81	1	2	2	3	3	2
40	60	65	70	74	79	85	90	64	68	73	78	84	91	1	2	1	1	1	-1
40	70	74	78	83	87	91	95	74	78	83	88	94	101	0	0	0	-1	-3	-6
40	80	83	87	90	93	96	98	84	88	93	98	104	111	-1	-1	-3	-5	-8	-13
50	20	23	26	29	31	34	37	23	27	31	35	39	45	0	-1	-2	-4	-5	-8
50	30	34	37	41	44	48	52	33	37	41	45	49	55	1	0	1	-1	-1	-3
50	40	44	48	52	56	61	66	43	47	51	55	59	65	1	1	2	2	2	1
50	50	54	58	62	67	72	77	53	57	61	65	69	75	1	1	2	3	3	2
50	60	64	68	72	76	81	86	63	67	71	75	79	85	1	1	2	2	2	1
50	70	74	77	81	84	88	93	73	77	81	85	89	95	1	0	1	-1	-1	-2
50	80	83	86	89	91	94	97	83	87	91	95	99	105	0	-1	-2	-4	-5	-8
60	20	22	24	26	29	30	32	23	26	29	32	36	40	-1	-2	-3	-3	-6	-8
60	30	33	36	38	41	44	47	33	36	39	42	46	50	0	0	-1	-1	-2	-3
60	40	43	46	50	53	56	60	43	46	49	52	56	60	0	0	1	1	0	0
60	50	53	57	60	64	68	72	53	56	59	62	66	70	0	1	1	2	2	2
60	60	63	67	70	74	78	82	63	66	69	72	76	80	0	1	1	2	2	2
60	70	73	76	79	82	86	90	73	76	79	82	86	90	0	0	0	0	0	0
60	80	82	85	87	90	93	96	83	86	89	92	96	100	-1	-1	-2	-2	-3	-4
70	20	22	23	25	26	27	28	22	25	27	29	33	36	0	-2	-2	-3	-6	-8
70	30	32	34	36	38	40	42	32	35	37	40	43	46	0	-1	-1	-2	-3	-4
70	40	43	45	47	50	52	54	42	45	47	50	53	56	1	1	0	1	-1	-2
70	50	53	55	58	61	64	67	52	55	57	60	63	66	1	1	1	1	1	1
70	60	63	65	68	71	74	78	62	65	67	70	73	76	1	1	1	2	1	2
70	70	72	75	77	80	83	87	72	75	77	80	83	86	0	1	0	1	0	1
70	80	82	84	86	89	91	94	82	85	87	90	93	96	0	-1	-1	-1	-2	-2

Decimal points omitted

Table 4
"Exact" hit rates for continuous criteria, HR1, and differences

		Exact						HR1						Differences					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	23	27	31	34	39	43	25	30	35	40	46	51	-2	-3	-4	-6	-7	-8
30	30	34	38	43	47	52	58	35	40	45	50	56	61	-1	-2	-2	-3	-4	-3
30	40	44	49	53	59	64	69	45	50	55	60	66	71	-1	-1	-2	-1	-2	-2
30	50	55	60	63	68	73	78	55	60	65	70	76	81	0	0	-2	-2	-3	-3
30	60	64	68	74	78	82	86	65	70	75	80	86	91	-1	-2	-1	-2	-4	-5
30	70	75	77	82	86	89	93	75	80	85	90	96	101	0	-3	-3	-4	-7	-8
30	80	83	85	88	91	93	96	85	90	95	100	105	111	-2	-5	-7	-9	-12	-15
40	20	23	26	29	32	35	38	24	28	32	36	40	45	-1	-2	-3	-4	-5	-7
40	30	34	37	41	44	48	52	34	38	42	46	50	55	0	-1	-1	-2	-2	-3
40	40	44	48	51	56	60	64	44	48	52	56	60	65	0	0	-1	0	0	-1
40	50	54	58	61	66	70	75	54	58	62	66	70	75	0	0	-1	0	0	0
40	60	63	68	71	75	80	83	64	68	72	76	80	85	-1	0	-1	-1	0	-2
40	70	74	77	81	84	86	89	74	78	82	86	90	95	0	-1	-1	-2	-4	-6
40	80	82	86	88	90	92	96	84	88	92	96	100	104	-2	-2	-4	-6	-8	-8
50	20	22	24	27	29	31	34	23	27	30	33	37	40	-1	-3	-3	-4	-6	-6
50	30	33	35	38	41	44	47	33	37	40	43	47	50	0	-2	-2	-2	-3	-3
50	40	43	46	50	53	56	60	43	47	50	53	57	60	0	-1	0	0	-1	0
50	50	53	56	60	63	67	70	53	57	60	63	67	70	0	-1	0	0	0	0
50	60	64	66	70	73	76	79	63	67	70	73	77	80	1	-1	0	0	-1	-1
50	70	73	76	78	81	84	87	73	77	80	83	87	90	0	-1	-2	-2	-3	-3
50	0	82	85	86	90	91	94	83	87	90	93	97	100	-1	-2	-4	-3	-6	-6
60	20	22	24	25	27	29	30	23	25	28	31	34	36	-1	-1	-3	-4	-5	-6
60	30	32	35	37	39	41	44	33	35	38	41	44	46	-1	0	-1	-2	-3	-2
60	40	43	45	47	50	53	55	43	45	48	51	54	56	0	0	-1	-1	-1	-1
60	50	52	55	58	61	63	67	53	55	58	61	64	66	-1	0	0	0	-1	1
60	60	63	65	68	70	73	76	63	65	68	71	74	76	0	0	0	-1	-1	0
60	70	72	75	77	79	82	85	73	75	78	81	84	86	-1	0	-1	-2	-2	-1
60	80	81	84	85	88	89	92	83	85	88	91	94	96	-2	-1	-3	-3	-5	-4
70	20	21	23	24	25	26	27	22	24	26	29	31	33	-1	-1	-2	-4	-5	-6
70	30	32	33	35	36	38	39	32	34	37	39	41	43	0	-1	-2	-3	-3	-4
70	40	42	44	46	47	50	51	42	44	47	49	51	53	0	0	-1	-2	-1	-2
70	50	52	54	56	58	60	62	52	54	57	59	61	63	0	0	-1	-1	-1	-1
70	60	62	64	66	68	70	73	62	64	67	69	71	73	0	0	-1	-1	-1	0
70	70	72	74	75	77	80	82	72	74	77	79	81	83	0	0	-2	-2	-1	-1
70	80	81	83	85	86	88	90	82	84	87	89	91	93	-1	-1	-2	-3	-3	-3

Decimal points omitted.

Table 5
"Exact" hit rates for continuous criteria, HR2, and differences

		Exact						HR2						Differences					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	23	27	31	34	39	43	25	30	36	42	49	58	-2	-3	-5	-8	-10	-15
30	30	34	38	43	47	52	58	35	40	46	52	59	68	-1	-2	-3	-5	-7	-10
30	40	44	49	53	59	64	69	45	50	56	62	69	78	-1	-1	-3	-3	-5	-9
30	50	55	60	63	68	73	78	55	60	66	72	79	88	0	0	-3	-4	-6	-10
30	60	64	68	74	78	82	86	65	70	76	82	89	98	-1	-2	-2	-4	-7	-12
30	70	75	77	82	86	89	93	75	80	86	92	99	108	0	-3	-4	-6	-10	-15
30	80	83	85	88	91	93	96	85	90	96	102	109	118	-2	-5	-8	-11	-16	-22
40	20	23	26	29	32	35	38	24	28	33	38	44	51	-1	-2	-4	-6	-9	-13
40	30	34	37	41	44	48	52	34	38	43	48	54	61	0	-1	-2	-4	-6	-9
40	40	44	48	51	56	60	64	44	48	53	58	64	71	0	0	-2	-2	-4	-7
40	50	54	58	61	66	70	75	54	58	63	68	74	81	0	0	-2	-2	-4	-6
40	60	63	68	71	75	80	83	64	68	73	78	84	91	-1	0	-2	-3	-4	-8
40	70	74	77	81	84	86	89	74	78	83	88	94	101	0	-1	-2	-4	-8	-12
40	80	82	86	88	90	92	96	84	88	93	98	104	111	-2	-2	-5	-8	-12	-15
50	20	22	24	27	29	31	34	23	27	31	35	39	45	-1	-3	-4	-6	-8	-11
50	30	33	35	38	41	44	47	33	37	41	45	49	55	0	-2	-3	-4	-5	-8
50	40	43	46	50	53	56	60	43	47	51	55	59	65	0	-1	-1	-2	-3	-5
50	50	53	56	60	63	67	70	53	57	61	65	69	75	0	-1	-1	-2	-2	-5
50	60	64	66	70	73	76	79	63	67	71	75	79	85	1	-1	-1	-2	-3	-6
50	70	73	76	78	81	84	87	73	77	81	85	89	95	0	-1	-3	-4	-5	-8
50	80	82	85	86	90	91	94	83	87	91	95	99	105	-1	-2	-5	-5	-8	-11
60	20	22	24	25	27	29	30	23	26	29	32	36	40	-1	-2	-4	-5	-7	-10
60	30	32	35	37	39	41	44	33	36	39	42	46	50	-1	-1	-2	-3	-5	-6
60	40	43	45	47	50	53	55	43	46	49	52	56	60	0	-1	-2	-2	-3	-5
60	50	52	55	58	61	63	67	53	56	59	62	66	70	-1	-1	-1	-1	-3	-3
60	60	63	65	68	70	73	76	63	66	69	72	76	80	0	-1	-1	-2	-3	-4
60	70	72	75	77	79	82	85	73	76	79	82	86	90	-1	-1	-2	-3	-4	-5
60	80	81	84	85	88	89	92	83	86	89	92	96	100	-2	-2	-4	-4	-7	-8
70	20	21	23	24	25	26	27	22	25	27	29	33	36	-1	-2	-3	-4	-7	-9
70	30	32	33	35	36	38	39	32	35	37	40	43	46	0	-2	-2	-4	-5	-7
70	40	42	44	46	47	50	51	42	45	47	50	53	56	0	-1	-1	-3	-3	-5
70	50	52	54	56	58	60	62	52	55	57	60	63	66	0	-1	-1	-2	-3	-4
70	60	62	64	66	68	70	73	62	65	67	70	73	76	0	-1	-1	-2	-3	-3
70	70	72	74	75	77	80	82	72	75	77	80	83	86	0	-1	-2	-3	-3	-4
70	80	81	83	85	86	88	90	82	85	87	90	93	96	-1	-2	-2	-4	-5	-6

Decimal points omitted.

the Taylor Russell tables, with a modal discrepancy near -.02, which increases to -.03 for HR2. Thus, the simple HR1 emerges as a superior compromise overall for approximating hr's for both binary and continuous criteria over the targeted parameter region (4). As the tables also show, beyond these ranges both approxi-mations deteriorate quickly.

5. Base Rate Problems

In loc.cit. we also briefly addressed the "base rate problem", which asks how much a valid test improves over random admissions from the unselected population of applicants in terms of the total percent of correct classifications,

(9) %c:=tp+tn

(cf. Table 1). To be sure, this is by no means the only plausible optimality criterion for evaluating the merit of a test. However, as Meehl and Rosen (1955) have stressed, it certainly warrants more attention than it has received in the past.

Especially for populations with severely skewed base rates - as they arise, for example, in clinical psychology - use of a test may be worse than no test at all when its use raises rather than lowers the proportion of misclassifications overall.

As a concrete illustration, Meehl and Rosen (1955) present a joint probability table (p. 198).The base rate of one of the two outcomes ("lower back pain is of organic origin") is .90, the quota is .66, and the probability of true positives is .63. For these figures, the validity of the test is between .2 and .3, which is not atypical for clinical tests. For the hit rate, one finds hr=.63/.90=.70, and for the success rate .63/.66=.95, all of which, so far, looks quite innocuous.

However, in terms of total percent correct, one finds from the implied joint probability table that %c=.63+.07=.70, which falls short of the base rate b=.90 for organic. Thus, if we throw away the test and diagnose all cases as "organic", only 1 out of 10 diagnoses will be incorrect. If we base the decision on the (valid) test, 3 out of 10 will be incorrect (Note 6).

This example, though contrived, is instructive since most predictive validities of commercial tests are actually in this range, at least for long range criteria worth predicting. For example, for college graduation, the SAT validities are near .2 (Crouse and Trusheim, 1988, p. 48), as are those for 8th semester college GPA (Humphreys, 1968). Thus, one might think that Meehl and Rosen's thought provoking discussion of the base rate problem would have stimulated much excitement in testing circles.

Actually, not much has changed since they wistfully observed that "Base-rates are virtually never reported" (Meehl and Rosen, 1955, p. 194). What did change is that it has become increasingly more difficult to locate predictive validities which have not been "corrected" upwards in some ingenious way (Note 7).

Table 6
Total percent correct for binary criteria and improvement (+) or deterioration (-) over random admission

		Total Percent Correct (%c)						%c - max (b, 1–b)					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	65	67	70	74	78	82	-5	-3	0	4	8	12
30	30	61	65	68	72	77	82	-9	-5	-2	2	7	12
30	40	58	61	65	70	74	79	-12	-9	-5	0	4	9
30	50	54	57	61	65	69	73	-16	-13	-9	-5	-1	3
30	60	50	53	56	60	63	66	-20	-17	-14	-10	-7	-4
30	70	45	48	51	53	56	58	-25	-22	-19	-17	-14	-12
30	80	40	43	45	46	48	49	-30	-27	-25	-24	-22	-21
40	20	59	62	65	68	71	74	-1	2	5	8	11	14
40	30	57	61	65	69	73	78	-3	1	5	9	13	18
40	40	56	60	64	68	73	78	-4	0	4	8	13	18
40	50	54	58	62	67	72	76	-6	-2	2	7	12	16
40	60	52	56	59	63	68	72	-8	-4	-1	3	8	12
40	70	49	52	56	60	63	66	-11	-8	-4	0	3	6
40	80	46	50	52	54	57	58	-14	-10	-8	-6	-3	-2
50	20	53	56	59	61	64	67	3	6	9	11	14	17
50	30	54	57	61	64	68	72	4	7	11	14	18	22
50	40	54	58	62	66	71	76	4	8	12	16	21	26
50	50	54	58	62	67	72	77	4	8	12	17	22	27
50	60	54	58	62	66	71	76	4	8	12	16	21	26
50	70	54	57	61	64	68	73	4	7	11	14	18	23
50	80	53	56	59	61	64	67	3	6	9	11	14	17
60	20	46	49	51	55	56	58	-14	-11	-9	-5	-4	-2
60	30	50	53	56	59	63	66	-10	-7	-4	-1	3	6
60	40	52	55	60	64	67	72	-8	-5	0	4	7	12
60	50	54	58	62	67	72	76	-6	-2	2	7	12	16
60	60	56	60	64	69	74	78	-4	0	4	9	14	18
60	70	58	61	65	68	73	78	-2	1	5	8	13	18
60	80	58	62	64	68	72	75	-2	2	4	8	12	15
70	20	41	42	45	46	48	49	-29	-28	-25	-24	-22	-21
70	30	45	48	50	53	56	59	-25	-22	-20	-17	-14	-11
70	40	50	53	56	60	63	66	-20	-17	-14	-10	-7	-4
70	50	54	57	61	65	70	74	-16	-13	-9	-5	0	4
70	60	58	61	65	69	74	79	-12	-9	-5	-1	4	9
70	70	61	65	68	72	76	82	-9	-5	-2	2	6	12
70	80	65	68	70	75	77	82	-5	-2	0	5	7	12

Decimal points omitted. If %c-max(b,1–b) positive, test better than random admissions, otherwise worse (in terms of %c).

Table 6 lists the total percent of correct classifications (%c) for binary criteria as a function of b, q, and r_{pb}. These values were obtained as a byproduct of the computations of the exact hit rates described earlier. As expected, they increase with r_{pb}, but also with the

(10) degree of synchrony of b and $q:=(b-.5)(q-.5)$,

which is positive if b and q depart in the same direction from .5, and negative if they depart in opposite directions (so that they do not match).

The right hand portion of Table 6 compares %c with the probabilities of correct decisions based on the larger base rate alone. If this difference is positive, the test increases the percent of correct decisions overall. If it is negative, it leads to a deterioration by the amount stated in the table. Cursory inspection of Table 6 shows that, for the realistic validity range ($r \leq .5$), use of a test ensures an increase in correct decisions uniformly only for base rates near .5. The more they depart from a 50/50 split, the more the benefits of using the test diminish. For base rates outside the (.3, .7) range, no test near the modal validity .3 improves over random admissions, regardless which quota is used. Short of this range, its benefit depends on the degree of synchrony, which can only be maximized if the base rates are known to locate the cut-off for the appropriate quota. But they are not known. As Meehl and Rosen noted 40 years ago: "the chief reason for our ignorance of the base rates is nothing more subtle than our failure to compute them" (p. 213).

Analogous results for continuous criteria, derived from the success rates (tp/q) in the tables in (Taylor and Russell, 1939), are presented in Table 7. For continuous criteria, the validity r is measured by the tetrachoric correlation. The overall results are very similar to those just discussed for binary criteria.

Finally, in Appendix 2, approximate formulae are derived for estimating the validity cut-off beyond which use of a test improves over random admissions (and betting on the outcome with the modal base rate). The first part (Appendix 2A) follows Meehl and Rosen (1955) to deduce the critical values in terms of hr (A2.5). Depending on which base rate is larger, two critical values are derived. These results are then extended to cut-offs for validities by invoking the HR1 approximation. This leads to the (rough under-) estimates (A2.10) and (A2.11):

(11) $\%c \geq b \geq .5 \Leftrightarrow r \geq 6(1-q)(b-.5)$, $\%c \geq 1-b \geq .5 \Leftrightarrow r \geq 6q(.5-b)$.

These estimates should be taken with a grain of salt in view of the various approximations invoked along the way. In practice, inspection of the %c tables (Tables 6 and 7) should suffice to gauge the merit of a test in terms of %c.

However, even approximate values are an improvement over the present practice of continuing to ignore the base rate problem altogether. To jar our collective memory, I close with one more quote from Meehl and Rosen's Psychological Bulletin paper

Table 7
Total percent correct for continuous criteria and improvement (+) or deterioration (-) over random admission

		Total Percent Correct (%c)						%c - max (b, 1–b)					
b	q	r:.1	.2	.3	.4	.5	.6	.1	.2	.3	.4	.5	.6
30	20	64	66	69	70	73	76	-6	-4	-1	0	3	6
30	30	60	63	66	68	71	75	-10	-7	-4	-2	1	5
30	40	56	59	62	65	68	71	-14	-11	-8	-5	-2	1
30	50	53	56	58	61	64	67	-17	-14	-12	-9	-6	-3
30	60	48	51	54	57	59	62	-22	-19	-16	-13	-11	-8
30	70	45	46	49	52	53	56	-25	-24	-21	-18	-17	-14
30	80	40	41	43	45	46	48	-30	-29	-27	-25	-24	-22
40	20	58	61	63	66	68	70	-2	1	3	6	8	10
40	30	57	60	63	65	68	72	-3	0	3	5	8	12
40	40	55	58	61	65	68	71	-5	-2	1	5	8	11
40	50	53	56	59	63	66	70	-7	-4	-1	3	6	10
40	60	50	54	57	60	64	66	-10	-6	-3	0	4	6
40	70	49	52	55	57	59	61	-11	-8	-5	-3	-1	1
40	80	46	49	50	52	54	57	-14	-11	-10	-8	-6	-3
50	20	52	54	57	59	61	64	2	4	7	9	11	14
50	30	53	55	58	61	64	67	3	5	8	11	14	17
50	40	53	56	60	63	66	70	3	6	10	13	16	20
50	50	53	56	60	63	67	70	3	6	10	13	17	20
50	60	54	56	60	63	66	69	4	6	10	13	16	19
50	70	53	56	58	61	64	67	3	6	8	11	14	17
50	80	52	55	56	60	61	64	2	5	6	10	11	14
60	20	46	49	50	52	55	56	-14	-11	-10	-8	-5	-4
60	30	48	52	54	57	59	63	-12	-8	-6	-3	-1	3
60	40	52	54	56	60	64	66	-8	-6	-4	0	4	6
60	50	52	56	60	63	66	70	-8	-4	0	3	6	10
60	60	56	58	62	64	68	71	-4	-2	2	4	8	11
60	70	56	60	62	65	68	72	-4	0	2	5	8	12
60	80	57	61	62	66	67	70	-3	1	2	6	7	10
70	20	39	42	44	45	46	48	-31	-28	-26	-25	-24	-22
70	30	45	46	49	50	53	55	-25	-24	-21	-20	-17	-15
70	40	49	52	54	56	60	61	-21	-18	-16	-14	-10	-9
70	50	53	56	58	61	64	67	-17	-14	-12	-9	-6	-3
70	60	57	60	62	65	68	72	-13	-10	-8	-5	-2	2
70	70	61	64	65	68	72	75	-9	-6	-5	-2	2	5
70	80	63	66	69	70	73	76	-7	-4	-1	0	3	6

Decimal points omitted. If %c-max(b,1-b) positive, test better than random admissions, otherwise worse (in terms of %c).

which, regrettably, has lost none of its urgency more than 40 years after it was first published:

"From the above illustrations it can be seen that the psychologist in interpreting a test and in evaluating its effectiveness must be very much aware of the population and its subclasses and the base rates of the behavior or event with which he is dealing at any given time." (Meehl and Rosen, 1955, p.199).

NOTES

Note 1: National Collegiate Athletics Association. A national watch-dog organization monitoring sports activities, especially football and basketball, at US colleges and universities.

Note 2: Estes (1992, p. 278) believes that "Intelligence tests ... are excellent predictors in many domains, ranging from school to a wide variety of occupations". However, he wisely refrained from supplying any supporting evidence for this bold claim.

Note 3: "Exact" within rounding error, i.e., based on the appropriate model, in contrast to an approximation based on simplifying assumptions. However, cf. Note 4 below.

Note 4: As in SDT, this "exact" ex-pression for r_{pb} is still contingent on two assumptions before it can be computed from 2×2 contingency tables: (a) Normality is needed to compute the mean difference from the conditional probabilities, and (b) Homoscedasticity is needed to render r, b, q identifiable in a 2×2 table of joint probabilities with three degrees of freedom. While both assumptions are routinely made in SDT, neither can be taken for granted. If one is willing to make them, then the common within variance can be set to unity, because a change of scale cancels in the numerator and denominator of r_{pb}.

Note 5: Taylor and Russell (1939) report success ratios only to two decimal places. In this case, the appropriate validity coefficient is the tetrachoric correlation.

Note 6: A reviewer remarked that Meehl and Rosen's "lower back pain example ... is an extreme case. If I know the base rate of the disease is as large as 0.9, I would use my common sense instead of a lower validity test to screen patients". I completely agree. However, (a), I believe it is fair to assume that Meehl and Rosen intended their extreme example as an illustration of a more general point; (b), Meehl was a clinical psychologist. In clinical psychology, base rates are frequently marginal; and (c), the continuing neglect of the base rate problem suggests that common sense was never a dominant factor in mental testing.

Note 7: It has become widespread practice to "correct" validities for attenuation. While such up-ward corrections may occasionally be defensible in theoretical work, it is less clear what purpose they are supposed to serve in applied work, when the test at hand does in fact contain measurement error weakening its predictive accuracy. Similarly, restriction of range corrections can only be justified, if at all, when based on the standard deviation of the subset of people actually taking the test (which may already be restricted through self selection), not on the standard deviation of the unselected population. In contrast, downward corrections of validity coefficients, for example corrections of multiple correlations for shrinkage, have become virtually extinct.

In the long run, it might help modulate the rising distrust of mental testing if such "corrected" validities were always accompanied by the original values, so that readers can draw their own con-clusions.

References

Cole, N. S. (1973) Bias in selection. *Journal of Educational Measurement, 10*, 237-255.

Crouse, J. and Trusheim, D. (1988) *The Case Against The SAT*. Chicago Illinois: The University of Chicago Press.

Estes, W. K. E. (1992) Ability testing: Postscipt on ability tests, testing, and public policy. *Cognitive Science, 5*, 278.

Freeman, H. (1963) Introduction to Statistical Inference. New York: Addison-Wesley.

Hartigan, J. R. and Wigdor, A.K. (1989) *Fairness in Employment Testing: Validity Generalization, Minor Issues, and the General Aptitude Battery*. Washington D.C., National Academy Press.

Humphreys, L. G. (1968) The fleeting nature of the prediction of college academic success. *Journal of Educational Measurement, 59*, 375-380.

Meehl, P. E. and Rosen, A. (1955) Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52*, 194-216.

Schönemann, P. H. and Thompson, W. W. (1996) Hit rate bias in mental testing. *Cahiers de Psychologie Cognitive/ Current Psychology of Cognition, 15*, 3-28.

Taylor, H. C. and Russell, J. T. (1939) The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

初稿收件：1997年12月1日
審稿通過：1998年7月16日

Appendix 1

A. Derivation of Exact Point Biserial Correlation

In Schönemann and Thompson (1996), we derived a simple approximation for hit rates,

(A1.1) $$HR1=q+r\sqrt{(1-b)/b'}/3,$$

after linearizing the cumulative normal distribution N(z) with

(A1.2) $$p:=N(z)\approx .5+z/3, \text{ with inverse } z\approx 3(p-.5).$$

This approximation was substituted into the formula for the point biserial correlation (here denoted "r" for simplicity):

(A1.3) $r'=d'\sqrt{b(1-b)}$, where

(A1.4) $d':=(m_Q-m_U)/s_w=m_Q-m_U$

is the standardized mean difference familiar from signal detection theory (SDT). I.e., m_Q denotes the mean of the qualified group, m_U that of the unqualified group, and s_w denotes the common within group standard deviation. It can be set to 1, since a change of scale affects the numerator and denominator in the middle of (A1.4) equally. This leads to

(A1.5) $r'\approx(m_Q-m_U)\sqrt{b(1-b)}$.

This approximation to r was used to derive the hit rate approximation HR1 in (A1.1). In loc. cit., we showed that this approximation works well within the parameter ranges

(A1.6) $(0<r<.5)$, $(.3<b<.7)$, $(.3<q>.7)$,

most likely to be relevant in validity studies.

However, strictly speaking, the derivations call for the total variance - not the common *within* group variance as in SDT - in the denominator of the squared mean difference, (A1.4). In general, this variance is given by

(A1.7) $s_t^2=bs_Q^2+(1-b)s_U^2+b(m_Q-m_t)^2+(1-b)(m_U-m_t)^2$,

where "m_t" denotes the total mean (Cf. Freeman, 1963, p. 185).

To be able to indentify r in a 2×2 joint probability table (which has 3 degrees of freedom), it is necessary to impose the homoscedasticity assumption ($s_Q=s_U$) familiar from ANOVA and SDT (Note 4). Under this constraint, (A1.7) reduces to

(A1.8) $s_t^2=b(1-b)(m_Q-m_U)^2+1$.

Thus, if one defines

(A1.9) $d''=(m_Q-m_U)/s_t$,

then the corrected expression for the point biserial becomes

(A1.10) $r=d''\sqrt{b(1-b)}$.

After making these substitutions, one finally arrives at

(A1.11) $r^2=(m_Q-m_U)^2b(1-b)/[(m_Q-m_U)^2b(1-b)+1]$
$=r'^2/(r'^2+1)$,

with r' as in (A1.5). On comparing the "corrected" expression r with the earlier approximation r' in (A1.4), one finds that they relate by the monotone transformation

(A1.12) $f(x)=\sqrt{x^2/(1+x^2)}$,

which has virtually no effect for small r'. For example, for $r'\approx.3$ one one obtains $r\approx.287$, and for $r'\approx.4$ one finds $r\approx.372$. In short, the erroneous denominator in (A1.4) is benign in the practically relevant validity range (A1.6).

B. Implicit Solution for Hit Rates

From the definitions set out in the joint probability table in Table 1, one finds that the false alarm rate can be expressed as a function of the hit rate, the base rate, and the quota:

(A1.13) $fr=(q-b.hr)/(1-b)$.

From the conditional probability table in Table 1, one further finds that the test cut-off C has respective representation

(A1.14) $C=m_Q+N^{-1}(1-hr)=m_U+N^{-1}(1-fr)$,

in the two within group distributions, so that

(A1.15) $m_Q-m_U=N^{-1}(1-fr)-N^{-1}(1-hr)$,

with fr as in (A1.13) and N( ) denoting the normal distribution function.

On substituting (A1.13), (A1.15) into (A1.11), one obtains an implicit expression relating b, q, r, with hr, which can be used to solve for hr iteratively. Table 2 and 3 give the resulting "exact" values for the hit rates as a function of the r_{pb}, b, and q. To obtain these values, initial hit rate estimates were successively refined until the discrepancy between argument r and reproduced r fell below .002. As already noted, these "exact" values are subject to the homoscedasticity constraint imposed on (A1.7) needed to achieve identifiability of 3 parameters in a 2×2 joint probability table, and the assumption that both within group distributions are normal.

C. Hit Rate Approximation Based on Corrected Variance

On invoking the linearizing approxi-mation (A1.2), the mean difference (A1.15) simplifies to

(A1.16) $m_Q - m_U = 3(fr - hr) = 3(hr - q)/(1 - b)$,

with a corresponding simplification of the point biserial:

(A1.17) $r \approx \sqrt{9(fr - hr)^2 b(1-b)/[9(fr - hr)^2 b(1-b) + 1]}$,

which, on invoking (A1.13), further reduces to

(A1.18) $r \approx \sqrt{[9(q - hr)^2 b]/[9(q - hr)^2 b + (1-b)]}$.

On solving this expression for hr, one obtains a new, "improved" approximation,

(A1.19) $HR2 := q + (r/\sqrt{1 - r^2})\sqrt{(1-b)/b}/3$,

which differs from HR1 only in the term $r/\sqrt{1 - r^2}$. Again, its effect is negligible for small r, which are the rule in most validity studies. Inspection shows that the "improved" formula leads to slightly larger estimates than HR1 as r grows larger. But in loc. cit. we found that HR1 already tends to overestimate the exact hit rates, so that the "improved" version HR2 further aggravates this bias (which, presumably, is induced by the linearization). Thus it turns out that the simpler approximation HR1, though derived from a flawed premise, is overall the better estimate. These observations are confirmed in Tables 2 through 5.

Appendix 2

Base Rates Versus Total Percent Correct Decisions

A. In Terms of Hit Rates

From the basic definitions of the joint probabilities on Table 1, one finds at once:

(A2.1) $\%c := tp + tn \geq b =: tp + fp \Leftrightarrow tn \geq fn$,

and also

(A2.2) $\%c := tp + tn \geq 1 - b =: fp + tn \Leftrightarrow tp \geq fp$

(see also the discussion in Meehl and Rosen, 1955, who, however, call %c "total hit rate"). Thus, to investigate under which conditions the total percent correct achieved with a test exceeds the total percent correct achieved without it -- by simply admitting randomly and betting on the outcome with the larger base rate, one has to distinguish between two cases: $b \geq .5$, and $b \leq .5$.

For the case $b \geq .5$, from the basic probability definitions in Table 1:

(A2.3) $$tp = b.hr,\ fn = b - b.hr,$$
$$tn = 1 - q - fn = 1 - q - b + b.hr.$$

Hence,

(A2.4) $$tn \geq fn \Leftrightarrow 1 - q - fn = 1 - q - b + b.hr \geq b - b.hr,$$

or

(A2.5) $$\%c \geq b \geq .5 \Leftrightarrow hr \geq 1 - (1 - q)/2b.$$

By the same logic one finds, for the case $b \leq .5$,

(A2.6) $$\%c \geq 1 - b \geq .5 \Leftrightarrow hr \geq q/2b.$$

B. In Terms of Correlations

It is natural to ask how the preceding results translate into lower limits for correlations necessary to lead to improvement in overall percent correct on using a (valid) test relative to random admissions (see also Tables 6 and 7 for explicit values). Approximate limits can be derived by use of the explicit hit rate estimates derived earlier. In view of the robustness results reported earlier, (cf. sec. 2 and Tables 2, 3, 4, and 5), the simpler hit rate approximation HR1 (A1.1) will be used. Again the two cases $b \geq .5$ and $b \leq .5$ will be considered separately.

For the case $b \geq .5$, substituting the expression for HR1 (A1.1) into (A2.5) gives

(A2.7) $$\%c \geq b \geq .5 \Leftrightarrow r \geq 3(1 - q)(2b - 1)/2\sqrt{b(1 - b)}.$$

Similarly, for the case $b \leq .5$,

(A2.8) $$\%c \geq 1 - b \geq .5 \Leftrightarrow r \geq 3q(1 - 2b)/2\sqrt{b(1 - b)}.$$

These approximations can be further simplified on noting that the multiplier

(A2.9) $$3|2b - 1|/2\sqrt{b(1 - b)} \approx 6|b - .5|.$$

For example, for $b = .4$ (or $b = .6$) one finds for the left side .61 and for the right side .6. If $b = .3$ (or $b = .7$), one finds 1.31 versus 1.2, which means, in practice, that use of (A2.9) slightly underestimates the minimum validity near the (.3, .7) bounds of the base rate range considered here.

Use of this simplification finally results in the rough bounds:

(A2.10) $\%c \geq b \geq .5 \Leftrightarrow r \geq 6(1-q)(b-.5)$

and

(A2.11) $\%c \geq 1-b \geq .5 \Leftrightarrow r \geq 6q(.5-b)$.

To illustrate, if b=.3 and q=.3, a validity of .36 is needed to exceed the percent correct obtainable with random admission alone (and predicting the outcome "unqualified" which arises with probability .7 in the unselected population). If the base rate is = .7 and the same admission quota q=.3 is retained, then the minimum validity necessary to improve over random admissions (and predicting that all admitted subjects are "qualified") rises to .84.

心理測驗命中率與基本率之新結果

舒彼得

國立台灣大學心理學系

本文爲最近有關命中率和基本率研究（Schönemann and Thompson, 1996）之延伸。文中修正了先前推導基本率近似値HR1時的錯誤前提，另提出稍較複雜之近似値HR2。然而在所探討的參數範圍內，HR1與HR2間僅具些微差異。本研究推導二分效標2×2列聯表的準確命中率，並將之與HR1，HR2，以及從泰羅二氏（1939）期望表經由貝氏定理爲連續效標推導之命中率三者比較。整體而言，較簡單之近似値HR1的表現優於HR2。最後，本研究以整體正確分類百分比爲考慮條件，推導出一個測驗若要比隨機甄選有效時，所需要的最低效度近似値。四十多年前，Meehl與Rosen（1955）即已提出警告：僅憑效度係數本身並不足以評斷一個測驗在應用上的優劣，因爲「當效標分類的基本率與百分之五十偏離甚大時，使用一個具備中度或低度效度的測驗，將會導致錯誤臨床決定的增加」（原著215頁）。本研究結果呼應了此點關切。

關鍵詞：心理測驗、智力測驗、命中率偏誤、基本率問題、預測效度