A COMPARATIVE STUDY OF SCORING ERRORS IN COUNTING NUMBER OF OMISSIONS AND RIGHT ANSWERS AND ONE AND TWO SIDED ANSWER-SHEET PRINTS

K. P. BHATTACHARYYA

Indian Statistical Institute

Two hypotheses are tested in the present study (1) whether scoring error in counting the number of right answers and omissions increases when the answer-sheets are printed on both the sides of a page and (2) whether scoring error is greater in counting the number of omissions (unanswered items) than in counting the number of right answers, enumerated with the aid of a scoring key.

Five psychometric tests were administered on nearly two thousand candidates. Double-sided answer-sheets were used for the first four tests, making a coupling of two tests on a single sheet of paper. For the remaining one, answer-sheets were printed single-sided. The first four hypothesis was rejected on Chi-square test ($p > .30$), while for the second hypothesis it was found that scoring error in counting the number of omissions was significantly greater ($p < .01$) than that of the right answers. Only in one case p was less than .05. Some improvements are suggested for minimising the scoring error when the answer-sheets are hand-scored.

Phillips and Weathers carried on a research investigation on the types of errors common on hand-scoring of objective type of standardized tests and found that 44.8 percent of the total scoring errors is contributed by the wrong counting of scores (1). The other types of errors (with their contributing percentages in parentheses) are as follows : instructions (26.1), use of key (14.9), use of tables (13.5) and computation (0.7)

In the present study two hypotheses are tested :

1. Whether scoring error in counting number of right answers and omissions increases when the answer sheets are printed on both the sides of a page and

2. Whether scoring error is greater in counting the number of omissions (unanswered items) than in counting the number of right answers, enumerated with the aid of a scoring key.

In the present case five objective type, psychometric tests were administered on nearly two thousand graduates for selecting candidates for a managerial course of studies with separate answer sheets. The tests were as follows :

Test No.	Name of the Test	Total No. of Items
1.1	General Ability	60
1.2	Graph & Table Reading	29
2.1	Breadth of Knowledge	38
2.2	English Knowledge & Comprehension	66
3.0	Mathematical Comprehension	25

*The author is thankful to Dr. S. Chatterji and Dr. (Miss) Manjula Mukerjee for suggestion and the data was supplied by them.

Of these five tests, the anwser sheets for test nos, 1.1 and 1.2 were printed on the same paper, one on each side of the page. The same arrangement was maintained for test nos, 2.1 and 2.2, while the answer sheet for test no. 3.0 was printed on single sheets, only one side of the sheets being used.

For scoring facility, the answer sheets were first sorted out test-wise and within each test answer-sheets for twenty candidates, or part thereof, arranged according to their roll number were separately bunched together and each of these bunches formed the unit of work-load for an individual scorer. There were altogether nine scorers, all having an academic education not less than class X of a higher secondary school and with previous experience of scoring psychometric tests. As none of the scorers were found to be error-free and finding the individual variation in scoring was not one of the objects of the present investigation, that aspect is not presented in this paper.

Each answer sheet was first hand-scored with a scoring-key, where the right answer-choices for the items are perforated, and then the number of omissions were counted. It was then re-scored independently by another scorer. The discrepancies found between scoring and rescoring were finally verified by a third one. For the application of the correction for guessing formula, the counting of both the number of right answers and omissions was necessary in the present study. The total number of errors found in each test were further split in three categories:

(a) Number of cases where the errors in couting the omissions exceeded that in counting the right scores (i.e., where E (om) > E (R), indicated by '+');

(b) Number of cases where the errors of counting the right scores exceeded that in counting the omis, sions (i.e., where E (om) < E (R)s indicated by '—');

(c) Number of cases where the errors in counting the right scores equalled with that of counting the omissions (i. e., where E (om) = E (R), indicated by '=').

Table 1 shows the raw data, along with the percentage conversions, given in parentheses.

TABLE 1

Showing Total Error (TE), No Error (NE), E (om) > E (R), E(om) < E(R), and E(om)=E(R) with percentage values in parentheses

Test No.	E (om) > E (R) (+)	E (om) < E(R) (—)	E(om) = E (R) (=)	Total Error (TE)	No Error (NE)
1.1	53	25	9	87	2
	(60.92)	(28.74)	(10.34)	(100.00)	
1.2	39	23	14	76	13
	(51.31)	(30.27)	(18.42)	(100.00)	
2.1	33	26	10	69	18
	(47.83)	(37.68)	(14.49)	(100.00)	
2.2	45	13	16	74	13
	(60.81)	(17.57)	(21.62)	(100 00)	
3.0	37	17	12	66	19
	(56.06)	(25.76)	(18.18)	(100.00)	

From Table 1 it is clear that the number of errors due to omissions is invariably greater than that of counting the right answers in all the five tests. And whether this difference is statistically significant or not is tested by chi-square test and reported afterwards.

For testing our first hypothesis we will require Table 2, where TE and NE are averaged for tests 1.1 and 1.2 and also for test 2.1 and 2.2, for in these two cases answer sheets were printed on both the sides.

TABLE 2

Showing Total Error (TE) and No Error (NE) with percentage values in parentheses for both types of answer sheets, printed single side and both the sides (averages are shown)

Test No.	Answer-sheet print	Total Error (TE)	No Error (NE)	Total (TE+NE)	Average TE	Average NE	Average TE+NE
1.1 + 1.2	Both sides	163	15	178	81.50 (91.57)	7.50 (8.43)	89.00 (100.00)
2.1 + 2.2	Both sides	143	31	174	71.50 (82.18)	15.50 (17.82)	87.00 (100.00)
3.0	Single side	66	19	85	66.00 (77.65)	19.00 (22.35)	85.00 (100.00)

It is obvious from Table 2 that though Test 3.0 had answer sheets, printed single side, the percentage distribution of TE and NE resembles more with the second row, and there is least resemblance between the first and second row, inspite of their having the same type of answer sheet, i.e., printed both the sides of a page. To test whether these differnces are significant or not, three λ^2-tests were carried out, those between tests 1 (1.1 and 1.2 combined) 3,2 (2.1 and 2.2 combined) and 3, and 1 and 2 and the result is presented in Tatble 3.

TABLE 3

Showing λ^2-and p-values for three sets of λ^2-tests

λ^2 -Test Between	λ^2	df	P
Test 1 and Test 3	1.144	1	.30
Test 2 and Test 3	.128	1	.80
Test 1 and Test 2	.506	1	.50

In all the cases p is greater than 0.3, so there is no significant difference in counting errors in tests where the answer-sheet is printed single-or both-sides of a page.

Now comes our second hypothesis: whether errors committed in counting omissions (i.e., unanswered items) is greater than those of counting right answers (i.e., filled up spaces). Let us persue Table 4, where errors in counting omissions execeeding those in counting right answers, i. e., E(om) > E (R), and the errors in counting right answers exceeding those in counting omissions, i.e , $E(om) < E(R)$ are shown with their totals, neglecting the cases where errors for these two aspects are equal with each other. The percentage values are presented in parentheses.

TABLE 4

Showing the number of errors in counting omissions exceeding that in counting right answers, i. e., $E(om) > E(R)$, its vice versa, i. e., $E(om) < E(R)$, the total of these two types of errors with their percentage value in parentheses.

Test No.	$E(om) > E(R)$ (+)	$E(om) < E(R)$ (—)	Total of '+' and '—'
1.1	53 (67.95)	25 (32.05)	78 (100.00)
1.2	39 (62.90)	23 (37.10)	62 (100.00)
2.1	33 (55.93)	26 (44.07)	59 (100.00)
2.2	45 (77.59)	13 (22.41)	58 (100.00)
3.0	37 (68.52)	17 (31.48)	54 (100.00)

To test whether the difference lying between columns 2 and 3 for each row is significant, five chi-square tests were performed and the results are reported in Table 5.

TABLE 5

Showing λ^2-values, df, and p for five tests to find the significance of difference between $E(om) > E(R)$ and $E(om) < E(R)$

Test No,	λ^2	df	p (less than)
1.1	12.180	1	.01
1.2	6.150	1	.01
2.1	1.179	1	.15
2.2	29.355	1	.01
3.0	12.989	1	.01

As the direction for the differences to be found was already indicated, i.e., '+' should be greater than '—', it was an one-tailed test, and the p-values are read accordingly. Here all the values are significant beyond one Percent level, except for Test 2.1. The reason may be that it being the test of Breadth of Knowledge, where questions on current affairs are set, for answering of which no special achievement is required the candidates might have been tempted to answer most of the problems, thus leaving least of omissions. And as the number of omissions dwindle, there is less probability for committing errors when counted by the scorers. Whether this assumption is corroborated by fact is worth another investigation. On the other hand, the situation is quite different in the case of other four tests. So it may be commented that of the two types of counting errors discussed in this paper, error in counting the omissions i e., blank spaces, is significantly greater (at least for those tests for answering of which some sort of achievement is needed) than that for counting the right answers.

But looking from the psychological standpoint we may ask: what could be the reason lying behind this empirical finding? We may assume that counting the right answers, extracts tenacious attention as the eyesight moves from one scoring-hole to another on the surface of the scoring stencil. But the scorer becomes less alert when counting the blank spaces, strewn haphazardly, here and there, along with several check-marks, acting as distractors. Now the eyes move without any hindrance, comparatively free, for the blank spaces are not localized to specific places on the answer sheet. When this operation is done, the scoring stencil is completely removed from the top of the answer-sheet and all the answer sheet lies under the scorer's eyes. Secondly, just after counting the right answers, where full attention is solicited, the scorer usually undertakes to count the omissions, much relaxed, not attuned to the level of attention needed so long, and without taking any rest. To minimise errors in counting omissions. counting of right answer with the help of the scoring stencil should be undertaken only after the counting of omissions for an answer sheet is over and some rest period should be introduced in between these two types of countings.

References

1. Phillips, B.N. and Weathers, G. "Analysis of errors made in scoring standardized tests." *Educational and Psychological Measurement*, 1958, 15, 563-567.
2. Siegel, S. *Nonparametric Statistics for the Behavioral Sciences*. New York: Mc Graw Hill, 1956.