Effect of Practice on Test Score

MANJULA MUKERJEE

(*Calcutta*)

The objective of the study was to test whether the previous experience of a similar test had any effect on the scores of the test taken afterwards. The scores on three pairs of tests conducted over three different groups of persons were analysed and the null hypothesis stated below was tested.

There is no practice effect or in other words the means of the test scores of the persons who did not have any experience of similar test is equal to that of the persons who had some experience of similar test.

Design of the study

In each of the three cases there were two tests. Each pair of tests contained items dealing with the same subject-matter. The number of items as well as the time limit, were same for both the tests.

In each case the available group of students (who took the tests) was split up into two groups at random. One group of students took one of the test say test A in the morning and the other test say test B in the afternoon. The other group took test B in the morning and test A in the afternoon.

In each of the three cases, these two sets of data were compared by the method of analysis of variance in order to test the null hypothesis stated above. This design would make possible not only a study of the effects of practice upon the test but also the interaction (*i.e.*, the relative effect of practice upon tests) of practice and test.

The main requirements which need to be fulfilled in order to apply analysis of variance are listed below :

(1) The effects of tests, practice and individuals upon the test scores are additive.

(2) The respective groups should be a simple random sample from a specified population. After the tests had been administered the test scores for each group might be regarded as random sample from the hypothetical population of test scores.

(3) The effects of interaction between persons and tests were normally and independently distributed with the same variances for each test and practice.

There was no way of testing whether condition (1) was satisfied in our case.

The persons who took the tests were not random sample from any specified population. We randomised available students into two groups. So under such a condition the population with regard to which inference might be drawn must be similar to the present sample. Lindquist (2) has shown that in most of the psychological experiments condition (3) is not satisfied and has stated that if any one of the conditions is only approximately satisfied then F will appear to be more significant than it really is.

Analysis of the scores

For all the three samples three pairs of tests were used, each pair of test dealing with different subject-matters. In the first case two tests on Mathematics knowledge were used. The second pair of tests were two forms of reasoning tests named Series Completion. In the third case we had two tests on computation. Three different groups of persons took these three pairs of tests. Results obtained as to the effect of practice on the test scores are presented below for each group separately.

TABLE 1

Analysis of variance (Mathematics Knowledge Tests)

Source of variation (1)	Degrees of freedom (2)	Sums of squares (3)	Mean sums of squares (4)	F (5)
Between Person	165	4380.952	26.551	1.040
Test × Practice	1	192.052	192.052	7.519**
Error (between person)	164	4188.900	25.542	...
Within Person	166	940.000	5.662	1.207
Test	1	125.349	125.349	26.721**
Practice ...	1	28.927	28.927	6.166*
Error (within person)	164	785.724	4.691	...
Total ...	331	5320.952	...	...

*Significant at 5% level. **Significant at 1% level.

The results showed that practice effect was significant at 5% but not at 1% level. So no immediate conclusion could be drawn from this. But here we would like to mention that the tests were too difficult for this particular

group concerned as shown by the large number of omission and low mean score and it might be true that practice effect was not pronounced only due to this fact. It was also found by applying t-test that for the test which was more easy than the other, the practice effect was significant whereas for the more difficult test it was not so.

TABLE 2

Analysis of variance (Series Completion Tests)

Sources of variations (1)	Degrees of freedom (2)	Sums of squares (3)	Mean sums of squares (4)	F (5)
Between Person	49	1943.360	39.660	2.064
Test × Practice	1	1020.704	1020.704	53.100**
Error (between person)	49	922.656	19.222	...
Within Person	50	1108.000	22.160	2.183**
Test	1	5.760	5.760	...
Practice	1	615.040	615.040	60.595**
Error (within person)	48	487.200	10.150	
Total	99	3051.000	...	

**Significant at 1% level.

Here practice effect, and interaction between practice and test were significant even at 1% level which proved that there was significant practice effect in this case. But here also, like Mathematics Knowledge Test the practice effect was more significant for the easier test (proved by t-test).

Analysis of computation test scores

Here the design was a bit different from the other two cases. There were two tests A and B with equal number of items dealing with the same subject. The available group of students was divided into two groups at random. One group took test A in the morning and took test BB' which contained all the items of B but the items were arranged in different order. The other group took test B in the morning and AA' in the afternoon where AA' contained all the items of test A but the items were arranged in different order. It should be remembered that we changed the order of items of test B to get test BB' in the

same way as we got test AA from test A. Here this design was adopted to test the effect of practice on the tests simultaneously with the effect of item arrangement. But it is to be noted that here the practice effect and the effect of item order were merged together. Hence only by neglecting the effect of item order we could analyse the scores in the same way as above. It is difficult to say how far this could be justified but under the circumstances nothing better could be done.

TABLE 3

Analysis of Variance (Computation Test)

Sources of variation (1)	Degrees of freedom (2)	Sums of squares (3)	Mean sums of squares (4)	F (5)
Between Person	115	8343.069	72.548	1.046
Test × Practice	1	395.173	395.175	5.668*
Error (between person)	114	7947.896	69.718	...
Within Person	116	1908.000	16.448	1.217
Test	1	19.931	19.931	1.475
Practice	1	347.655	347.655	25.730**
Error (within person)	114	1540.414	13.512	...
Total	231	10251.069	...	...

*Significant at 5% level. **Significant at 1% level.

From the obtained results we could say that in this case also there was some practice effect, as the practice effect was significant at 1% level. Here it should be noted that the interaction was not significant at 1% level. It should be remembered in this connection that if item arrangement in this type of test had any effect on test scores, it had also contributed its share towards making the result what it is.

Conclusions

Analysis of data did show in all the three samples studied that practice had some effect on test scores and that this effect was significant statistically. But as the data collected were not mainly for this experiment alone, the rigorous

controls which would have been necessary for doing this experiment could not be met.

The defects in our experiments have been mentioned earlier. With the first sample we did not get practice effect significant at 1% level ; it might be due to the fact that the tests were too difficult. In the last case we had to assume something of which we were not very much sure and also could not test it from our analysis.

Considering all these it could be concluded that though our analysis showed that there was some practice effect yet we could not totally reject our null hypothesis. Further study with adequate control is needed to test the issue. Results of the present study however strengthen the argument for conducting such a study because the obtained results suggest strongly that practice may have a significant effect on test scores.

Suggestions

We have already mentioned that in the last case the adopted design was not an ideal one. In order to test the effect of item order simultaneously with practice effect we could adopt any one of the following design (1, pp. 288-292).

(1) If we want to test the effect of item order by within error sum of square we should divide the available group of persons into four groups. One group should take test A in the morning and test BB' in the afternoon ; second group should take test B in the morning and test AA in the afternoon ; the third group should take test AA' in the morning and test B in the afternoon ; the last group should take test BB' in the morning and test A in the afternoon. The analysis of variance table will be of the following form (1, pp. 288-292).

Sources of variation	Degrees of freedom	Sums of squares	Mean sums of squares	F
Between Persons				
Test × Practice				
Test × Form				
Form × Practice				
Error (b)				
Within Persons				
Test				
Practice				
Form				

(2) If we think that the effect of item arrangement is not important then we can adopt the following design. As before the group of persons should be divided into four groups. The first group should take test A in the morning and test B in the afternoon ; the second group should take test B in the morning and test A in the afternoon ; the third group should take test AA' in the morning and test BB' in the afternoon ; the last group should take test BB' in the morning and test AA' in the afternoon. The analysis of variance table will be of the following form (1, pp. 285-288).

Sources of variation	Degree of freedom	Sums of squares	Mean sums of squares	F
Between Persons				
Form				
Test × Practice				
Test × Form × Practice				
Error (b)				
Within Persons				
Test				
Practice				
Form				

REFERENCE

1. LINDQUIST, E. F, (1953). Design and Analysis of Expe iments in Psychology and Education, Honghton Mifflin Company, Boston.