

EFFECT OF PRACTICE IN MARKING ANSWER SHEETS ON SUBSEQUENT TEST RESULTS

H. George Loiselle
Dade County

An experiment was conducted at the fifth-grade level to determine whether separate machine-scored answer sheets could be used without penalizing pupils' scores (Loiselle). The results indicated that the group which used the expendable booklet edition of the test received statistically significant greater scores. It was concluded that further studies should be made to determine whether a period of orientation which would involve several opportunities to record answers on separate machine-scored answer sheets prior to testing might eliminate the differences in scores between the two groups.

Procedures

A new experiment was undertaken at the fifth-grade level to determine whether practice in using answer sheets prior to testing would remove the differences noted above. By means of memoranda and a radio broadcast, the Supervisor of Testing instructed the principals and teachers in the method of administering the test and determining control and experimental groups within each school. Each teacher was instructed to select his control and experimental group by checking every other name on an alphabetical list of his class starting with the first or second name. The names with the check were designated as the experimental group. These pupils were administered the Stanford Achievement Test, Intermediate Battery Partial Form KM, and used separate machine-scored answer sheets. The other pupils became the control group and were administered the Stanford Achievement Test, Intermediate Battery Partial Form K, the consumable edition which had previously been selected as part of a county-wide testing program.

To eliminate possible variations in the administration and timing of the test, the Supervisor of Testing personally administered the tests over the school system's FM radio band to the control and the experimental groups. Approximately 9,000 pupils were tested in this way. Each principal designated those teachers who proctored the control groups and those who proctored the experimental groups. Each school was asked to utilize its public address system for a short orientation period to orient the pupils in following directions given by radio. Materials on the proper manner of using a separate answer sheet and recording answers were provided the schools for this purpose. Practice tests with separate answer sheets were provided to teachers for their experimental

pupils so that they could participate in four short testing sessions prior to the actual testing date. Control pupils received only general orientation on the manner of filling in the answer spaces without practice sessions.

Final test answer sheets were scored on IBM test scoring machines at the Testing Department and test booklets were scored by the teachers. Each school prepared data sheets, in alphabetical order, for each of the two groups within the school. Data sheets, answer sheets, and all scored consumable booklets were then returned to the Testing Department. The sample was obtained in the same manner as the previous year; every tenth booklet was drawn in order to obtain the control sample for the statistical study. Each consumable booklet thus selected was then re-scored by trained clerical workers who made out new data sheets for the county control sample. Data for the experimental group were obtained by selecting every tenth name on the alphabetical list of the experimental pupils within each school tested by radio; separate data sheets were also made out for this sampling of the experimental group. Raw scores were used in the statistical work.

Analysis

In the control sample, there were 481 cases and in the experimental there were 441. Means and standard deviations were computed for each of the groups in the six different areas of achievement measured by the test: paragraph meaning, word meaning, spelling, language, arithmetic reasoning, and arithmetic computation. The t-test for significant differences between groups was applied.

Table 1 shows that there was no significant difference between the control and the experimental groups in paragraph meaning, word meaning, language, arithmetic reasoning, and arithmetic computation; spelling was significant in favor of the control group at the .01 level.

When a composite test of significance as shown in Table 2 was applied, it was found that there was no significant difference between the two groups.

To determine whether a difference in mental ability between the control and experimental groups would cause a significant readjustment of the means in paragraph meaning, which approached significance in favor of the control group, and in language, which approached significance in favor of the experimental group, an analysis of covariance was made. In each instance, the results of this analysis indicated that there was no significant difference between the adjusted means of the two groups.

Table 1

Significance of the Differences in Six Areas of the
Stanford Achievement Test, Intermediate Battery Partial

Sub-Test	Control (N=481)		Experimental (N=441)		Difference in Means	t
	Mean	SD	Mean	SD		
Paragraph Meaning	24.05	9.93	23.22	9.37	+ .83*	+1.33
Word Meaning	24.39	11.07	24.72	10.70	- .33	- .47
Spelling	35.27	13.05	32.66	12.89	+2.71	+3.22**
Language	23.64	14.66	24.83	13.11	-1.19	-1.32
Arithmetic Reasoning	22.77	8.69	22.59	8.71	+ .18	+ .32
Arithmetic Computation	16.41	6.29	16.09	6.42	+ .32	+ .41

*A plus value denotes a difference in favor of the control group.

**Significant at .01 level of confidence.

Table 2

Composite Test of the Significance of the Differences
Obtained in the Six Areas Studied

	t	P	$-\log_e P$	df
t_1	+1.33	.1836	1.69500	2
t_2	- .47	-.6384	- .44942	2
t_3	+3.22	.0013742	6.59294	2
t_4	-1.32	-.1868	-1.67773	2
t_5	+ .32	.7490	.28902	2
t_6	+ .41	.6818	.38303	2
		Total	6.83284	12
			$X^2 = 2 (6.83284) = 13.66568$	with 12 df

Results and Conclusions

In paragraph meaning, word meaning, language, arithmetic reasoning, and arithmetic computation, there is no statistically significant difference between the means of the two groups.

In spelling, there is a statistically significant difference between the means of the two groups in favor of the control group which recorded its answers in the test booklet.

A composite test of significance, the chi square test, revealed no significant difference between the two groups.

It may be concluded from this experiment that when a period of orientation involving several opportunities to record answers on separate machine-scored answer sheets is given prior to the actual testing, the scores of fifth-grade pupils who take the machine-scored edition of the Stanford Achievement Intermediate Partial Battery, Form KM, are not significantly different (except in spelling) from those of the fifth-grade pupils who take the test using the expendable booklet edition.

The printing and supplying of separate practice tests and answer sheets to large groups of pupils is rather expensive and should be taken into consideration when making a decision whether or not to use separate machine-scored answer sheets.

Reference

Loiselle, H. George. An experiment in the use of machine-scored answer sheets with fifth-grade pupils. Florida Journal of Educational Research, 1961, III, 33-37.