

## COMMUNICATIONS AND DISCUSSIONS

---

### A CLASS EXPERIMENT WITH THE HILLEGAS SCALE

The following study was made in order to illustrate, for classes in educational psychology, the use of the Hillegas Scale.\* Twenty-five short compositions were taken as the material to be graded. These had all been written by college freshmen, and were all reproductions of the story of the "Marble Statue." The judges who did the grading were forty-one women, Junior and Senior students in educational psychology. They had had no experience in grading such compositions.

Each student had a copy of the twenty-five essays, and she was directed to grade them first without the use of any objective scale, giving to each any score from zero to one hundred which she felt that the composition deserved. A week later the use of a scale was explained, and these students were then asked to grade the same set of compositions by means of the Hillegas Scale.

Table I gives the results of the grading of the forty-one judges. In the first column is the number of each composition. Columns two, three and four give results without the scale, and columns five, six and seven give results with the scale. In columns two and five are given the variations in range of the judgments on each composition. In columns three and six are the medians of the judgments, and in columns four and seven are the standard deviations of the judgments on each composition.

An examination of this table shows the following:

1. The range of judgments is absolutely very wide, both with the scale and without it, but

2. The range is less with the scale. The average reduction is 11.5 points.

3. The range of the medians is the same, 32 points, with the scale and without it. Therefore, though the scale reduces the range of judgments on each composition, it does not do so at the cost of reducing the difference between the best and the worst compositions.

4. The standard deviations are reduced in the cases of all except two of the compositions. The average S. D. without the scale being 11.2 and the average S. D. with the scale 8.9.

These points would seem to show that the tendency of the scale is to make judgments more stable.

Another way of inquiring into the effect of the scale is to compute the correlation coefficients for various pairs of judges. It seems fair to assume that if the scale tends to promote uniformity it will

---

\*M. B. HILLEGAS. *A Scale for the Measurement of Quality in English Composition*. Teachers College Publications.

TABLE I

*Judgments of Forty-one Students upon Twenty-five Compositions*

WITHOUT SCALE				WITH SCALE		
I. Compo- sition no.	II. Size of Range	III. Median	IV. S. D.	V. Size of Range	VI. Median	VII. S. D.
1	45	65	11.1	41	67	9.0
2	50	65	13.0	35	67	9.7
3	55	75	10.3	48	76	9.4
4	47	85	10.7	40	77	9.8
5	58	68	12.8	47	65	10.2
6	42	75	11.2	37	72	8.6
7	50	75	14.5	38	72	9.6
8	45	68	11.0	33	67	8.8
9	50	70	11.8	38	68	9.1
10	50	80	9.6	38	75	8.6
11	23	90	5.9	22	85	6.4
12	70	75	12.3	36	78	8.8
13	45	80	8.9	32	75	8.5
14	60	74	10.8	41	73	9.1
15	66	58	16.9	35	55	9.9
16	70	60	16.0	48	67	9.9
17	35	80	9.6	32	77	8.1
18	53	75	11.4	45	75	9.8
19	62	69	12.7	40	69	9.3
20	45	80	10.3	37	75	8.1
21	70	70	16.0	55	68	11.5
22	30	85	8.0	36	83	7.1
23	45	70	8.9	34	72	9.4
24	39	80	8.7	34	77	7.8
25	30	85	7.7	25	87	6.9
Average 49.4				37.8		8.9

tend to raise the correlations between judges. In order to test this point the intercorrelations between forty pairs of judges were computed. For each pair the correlations were figured between their judgments before using the scale, and then between their judgments with the scale. In Table II the numbers in the first column stand for the various pairs of judges. The second and third columns give the Pearson coefficients as between judgments without the scale and with the scale respectively.

TABLE II  
*Correlations Between Various Pairs of Judges*

Pair No.	Without Scale r =	With Scale r =	Pair No.	Without Scale r =	With Scale r =
1	.11	.36	21	.51	.63
2	.25	.49	22	.51	.39
3	.25	.32	23	.52	.39
4	.26	.00	24	.52	.28
5	.29	.26	25	.54	.32
6	.29	.45	26	.54	.67
7	.32	.52	27	.56	.53
8	.35	.19	28	.57	.55
9	.36	.54	29	.57	.39
10	.38	.29	30	.58	.67
11	.41	.35	31	.59	.70
12	.41	.42	32	.59	.51
13	.41	.48	33	.60	.62
14	.43	.27	34	.61	.62
15	.46	.55	35	.62	.63
16	.46	.44	36	.63	.20
17	.47	.45	37	.66	.24
18	.47	.39	38	.66	.78
19	.48	.64	39	.75	.52
20	.50	.57	40	.86	.70
			Average	.48	.46

From this table it appears that the range of correlations without the scale is from .11 to .86, and with the scale is from .00 to .78. The average correlation is .48 without the scale and .46 with the scale. In order to compare the judgments of different groups the forty-one students were divided into two classes of twenty and twenty-one members. The average scores of the twenty-five compositions as judged by the first class were correlated with the average scores as judged by the second class. This was done for the class judgments without the scale and with the scale. By the Pearson formula  $r = .87$  in both cases. By the formula  $1 - \frac{6 \sum D^2}{n(n^2 - 1)}$  the figures were .92 without the scale and .87 with it.

At first thought these correlation results seem to be in opposition to the results as shown in the standard deviations. But I doubt whether there is a true contradiction here. It seems to me compatible to say that, whereas the scale does steady the judgment of a group with respect to its marking of a single composition, it does not alter the judgment, either of the individual or the group, with respect to the relative merits of a series of compositions.

KATE GORDON.

Carnegie Institute of Technology.