

Use of Ratings in the Evaluation of Exhibits

HOMER N. CALVER, F.A.P.H.A., MAYHEW DERRYBERRY,
PH.D., F.A.P.H.A., AND IVAN N. MENSCH

Secretary, American Museum of Health, New York, N. Y.; Chief, Field Activities in Health Education; and Junior Social Science Analyst, U. S. Public Health Service, Washington, D. C.

THE evaluation of exhibits in health education becomes increasingly essential as public health workers make more frequent use of that medium. At almost all state or county fairs, conventions, or other large gatherings there will be found one or more health exhibits, many of which are complicated and expensive. Despite increased interest in using exhibits, there is little objective evidence concerning their effectiveness. Generally the sole evaluation of an exhibit before it is put on display is the judgment of the exhibitor and his associates. Are such ratings dependable, and can large expenditures for exhibits be justified on subjective opinions?

That the validity of ratings may be open to some doubt is suggested by previous studies of ratings on the quality of drawings, English compositions, and handwriting. As early as 1910 Thorndike asked 30 judges to rate 37 specimens of handwriting for quality.¹ In rating one of the specimens, one judge stated that the writing was of the highest quality; 5 stated it was of the poorest; and the ratings by the other 24 judges were distributed throughout the range between best and poorest. Later, Thorndike,² Cohen,³ and others showed similar variations in ratings of drawings, and Hillegas⁴ demonstrated that

ratings are unreliable measures of quality in English composition.

Although as a result of such studies those in educational measurement no longer use ratings as a means of evaluating educational materials or products, constructors of health exhibits have seldom utilized any other measure of effectiveness. For that reason it has seemed advisable to investigate exhibit ratings to ascertain whether such evaluations are reliable; and the New York World's Fair presented an excellent opportunity for that investigation. Grouped in the Medicine and Public Health Building were many exhibits prepared by many health workers, with extremely diverse ideas as to effective exhibit technic, including all sorts of combinations of movement, color, and arrangement. Twenty-six major ones were selected for study.

In evaluating, two questions were posed: Could professional workers agree on the most effective exhibits in the group? Could they also agree in selecting the least effective displays?

RATINGS BY PROFESSIONAL WORKERS

Two hundred and seventy-one professionally trained persons, including public health workers, educators, students in professional school, and professional exhibitors were asked to

TABLE 1

Relative Frequency with Which 26 Health Exhibits Were Rated as "Best" or "Worst" by Professional Workers

Name of Exhibit	Percentage of Professional Workers Rating the Exhibit as One of the Three—	
	"Best"	"Worst"
Maternal Health	51.7	4.1
Maze of Superstition	26.9	10.3
Education of the Blind	25.8	1.1
Industrial Sanitation	24.7	0.7
Cancer	20.7	4.4
Dental Science	18.1	3.7
Pneumonia	17.7	8.5
Milk Control	15.5	3.7
Carrel-Lindbergh Apparatus	14.4	8.5
Heart and Circulation of Blood	12.6	8.1
Allergy	12.6	5.9
Glands of Internal Secretion	10.7	4.4
Immunity, Infections, Viruses	8.9	4.1
Syphilis—"Shadow on the Land"	6.6	5.9
Housing for Health	6.3	22.9
Tuberculosis	5.2	12.9
Chest X-ray Demonstration	4.8	17.7
Mental Hygiene	4.4	8.5
Social Hygiene in Your Town	3.7	3.3
Your Health Department	3.3	28.0
Medical Education	2.6	13.3
Safeguarding Medicinal Products	2.6	11.1
The Anemias, Story of Diabetes	2.2	13.3
Anesthesia	2.2	8.1
Veterinary Medicine	1.9	11.8
Child Health	0.7	16.2

review the exhibits and select the first, second, and third best as well as the three which they considered as the worst of the 26. Judgments were to be made in accordance with the following instructions:

Please rate the exhibits in the Hall of Medicine on the basis of *at least* the following characteristics: (1) clarity of presentation, (2) general attractiveness, (3) integration of panels, and (4) value of information portrayed.

In addition, the rating form provided space for the judge to state his reason(s) for rating any given exhibit as "best" or "worst."

A tabulation of the frequency with which each exhibit was rated as one of the three best or one of the three worst is shown in Table 1.

Low percentages appearing in both columns for a given exhibit indicate marked lack of agreement among the judges on its relative merit. On only one exhibit (*Maternal Health*) did as many as half the judges agree that it

was among the three best exhibits shown, and even on that one 11 of the 271 raters (or 4.1 per cent) considered it as one of the three worst.

No other exhibit was placed in the best or worst position by a majority of the raters; only 6 others were rated in a similar fashion by as many as one-fifth of the raters. For example, 73 judges rated the *Maze of Superstition* exhibit as one of the three best, but 28 others called it one of the three worst. Clearly, these data indicate that the criteria which the raters used were not applied consistently by the 271 individuals.

Further evidence of the lack of agreement between the judges is afforded through the following quotations from the rating blanks on given exhibits.

It may be argued that much of the disagreement may have been due to the different types of professional groups asked to rate the exhibits. To test that hypothesis, the judges were divided into four more or less homogeneous group-

QUOTATIONS FROM RATING BLANKS

Anesthesia Exhibit

"This exhibit portrays the place of anesthesia in modern surgery in such a manner that fear of possible pain and fear of anesthetic accident is removed. The scene of operation is quite accurately portrayed, with no attempt to dramatize the situation. The life-size moving figures are convincing. One leaves the exhibit with a feeling of reassurance."

"Here is an opportunity for a fine piece of work and what is there here—nothing but a hospital operating room—offers nothing to onlooker in an educational way. One glance and people walked away."

Dental Exhibit

"Presented very clearly and attractively; panels very well integrated. Affords definite information as to cause of dental decay and malformation impressing necessity of prenatal care, area (enamel) attacked first, correct method of cleaning teeth, value of early and regular visits to dentist; how personal appearance is improved by insertion of dentures. An excellent visual method of emphasizing care of teeth, especially for children."

1. Negative
2. Uninteresting
3. Rather gruesome
4. Poorly organized
5. No participation by observers
6. Exhibit lost a fine opportunity of presenting some fine lessons in practical dental health education."

Social Hygiene Exhibit

"The general idea of presentation I thought was very good as well as being attractive. It makes one conscious of what the community can do for one. The general placing of the institutions was instructional. Exhibits of this type would lead to better health (mental and physical) and a general improved community life. It showed good integration."

"The colors are dull and lack all visual appeal. The name social hygiene, I believe, should be explained to the lay public. It is a somewhat misleading term. Many illustrations are used as a method of conveying information. These are inadequate, due to poor workmanship and lack of sufficient explanation. The name 'social' leads to much confusion on the entire exhibit. The integration of panels is without any logical sequence."

ings: public health workers (112 in number), educators (84), students of medicine or education (66), and professional exhibitors (9); and their ratings tabulated separately. The percentage of agreement in each professional group is shown in Table 2. There is no evidence in that table to indicate that there is any closer agreement when the ratings are tabulated by homogeneous professional groups than when they are considered as though made by a mixed group of professional workers. In only one homogeneous group and for only

one exhibit (the students, in rating *Maternal Health*) was there a majority opinion on a specific exhibit. The same general lack of agreement exists throughout the table.

WHAT THE PUBLIC THINKS OF EXHIBITS

In order to determine whether the reactions of the general public (who, after all, are the ones for whom the exhibits were prepared) were more reliable than the judgments of experts, 18,449 lay spectators also were asked to rate the exhibits. Their instructions

TABLE 2

Relative Frequency with Which 26 Health Exhibits Were Rated as "Best" or "Worst" by 4 Professional Groups

Exhibit	112 Public Health Workers		84 Educators		66 Students of Medicine or Education		9 Professional Exhibitors	
	Best ¹	Worst ²	Best	Worst	Best	Worst	Best	Worst
Maternal Health	49.1	0.0	47.6	7.1	62.1	6.1	44.4	11.1
Maze of Superstition	38.4	0.9	19.0	1.2	7.6	0.0	33.3	0.0
Education of the Blind	33.0	1.8	16.7	0.0	27.3	1.5	11.1	0.0
Industrial Sanitation	27.7	2.7	10.7	4.8	21.2	7.6	22.2	0.0
Cancer	21.4	4.5	19.0	2.4	10.6	1.5	22.2	22.2
Dental Science	20.5	3.6	19.0	7.1	9.1	18.2	33.3	11.1
Pneumonia	14.3	13.4	41.7	2.4	28.8	15.2	33.3	11.1
Milk Control	14.3	3.6	16.7	3.6	16.7	4.5	11.1	0.0
Carrel-Lindbergh Apparatus	12.5	4.5	16.7	8.3	7.6	6.1	11.1	0.0
Heart and Circulation of Blood	9.8	8.9	16.7	7.1	16.7	10.6	33.3	0.0
Allergy	8.0	8.0	9.5	3.6	25.8	15.2	0.0	0.0
Glands of Internal Secretion	8.0	5.4	6.0	6.0	3.0	6.1	22.2	11.1
Immunity, Infections, Viruses	8.0	2.7	9.5	7.1	10.6	3.0	0.0	0.0
Syphilis—"Shadow on the Land"	7.1	16.1	4.8	4.8	1.5	16.7	11.1	22.2
Housing for Health	5.4	7.1	6.0	13.1	1.5	6.1	0.0	0.0
Tuberculosis	4.5	24.1	7.1	20.2	9.1	27.3	0.0	0.0
Chest X-Ray Demonstration	4.5	12.5	2.4	23.8	9.1	16.7	0.0	33.3
Mental Hygiene	4.5	11.6	2.4	16.7	0.0	1.5	0.0	22.2
Social Hygiene in Your Town	4.5	6.3	1.2	15.5	0.0	3.0	0.0	0.0
Your Health Department	4.5	2.7	8.3	6.0	25.8	6.1	0.0	0.0
Medical Education	4.5	2.7	6.0	6.0	0.0	1.5	0.0	0.0
Safeguarding Medicinal Products	2.7	33.9	4.8	19.0	1.5	28.8	11.1	33.3
The Anemias, Story of Diabetes	2.7	3.6	3.6	19.0	1.5	21.2	0.0	22.2
Anesthesia	0.9	17.9	1.2	16.7	0.0	15.2	0.0	0.0
Veterinary Medicine	0.9	9.8	2.4	10.7	1.5	16.7	11.1	11.1
Child Health	0.9	8.0	2.4	10.7	4.5	25.8	0.0	11.1

¹ Per cent of group rating exhibit as one of the three best.

² Per cent of group rating exhibit as one of the three worst.

TABLE 3

Relative Frequency with Which 26 Exhibits Were Rated as "Liked Most" or "Liked Least" by 18,449 Lay Spectators

Exhibit	Per cent Liked Most	Per cent Liked Least
Tuberculosis	28.1	8.6
Carrel-Lindbergh Apparatus	28.1	2.2
Industrial Sanitation	26.5	3.2
Education of the Blind	26.1	2.6
Safeguarding Medicinal Products	25.0	2.8
Cancer	24.0	3.5
Heart and Circulation of Blood	23.0	5.1
Social Hygiene in Your Town	21.9	5.0
Pneumonia	19.2	2.8
Housing for Health	18.6	5.8
Allergy	18.4	2.7
Anesthesia	15.0	3.8
Milk Control	13.8	5.5
Maternal Health	11.9	8.3
Glands of Internal Secretion	11.6	16.0
Chest X-Ray Demonstration	11.6	6.8
Your Health Department	11.5	3.9
The Anemias, Story of Diabetes	11.4	5.6
Maze of Superstition	9.7	6.7
Dental Science	9.4	3.8
Syphilis—"Shadow on the Land"	9.3	5.1
Immunity, Infections, Viruses	6.6	9.9
Child Health	6.5	9.1
Veterinary Medicine	6.2	21.1
Mental Hygiene	5.0	10.3
Medical Education	1.6	5.0

were to check, on a rating form, the exhibits at which they had stopped. On the form were listed the 26 major exhibits which had been rated by the professional workers, and beneath the list were the following instructions:

"Of the exhibits checked above, please list the three exhibits you liked most and the three you liked least."

Spaces were provided for the choices.

Tabulation of their preferences, as shown in Table 3, indicates no direct overwhelming agreements. For only 8 exhibits did more than one-fifth of the spectators express a preference.* It would seem that judgments by laymen are also unreliable as criteria to judge the effectiveness of exhibits.

DO PROFESSIONAL WORKERS AND THE PUBLIC AGREE?

There was also considerable disagreement between the ratings of the professional group and those of the laymen. Of the former, 51.7 per cent rated the *Maternal Health* exhibit as best, but only 11.9 per cent of the laymen gave it the same rating. Similarly, 16.0 per cent of the laymen rated *Glands of Internal Secretion* as worst, whereas only 4.4 per cent of the professionals gave it the same rating. In rating another exhibit (*Tuberculosis*) 28.1 per cent of the laymen indicated it as one of the best, yet only 5.2 per cent of the professional raters made a similar selection.

Thus, there was disagreement among the professional groups constructing, using, and evaluating exhibits; among the members of the public for whom the exhibits were intended; and between those making the exhibits and those for whom the exhibits were made. What, then, is the value of ratings of this type in the evaluation of exhibits?

SUMMARY AND CONCLUSIONS

The increased interest in exhibits and their use for health education warrants an evaluation of their effectiveness if expenditures for exhibits are to be justified. An attempt at such an evaluation of the health exhibits at the New York World's Fair was undertaken in 1939. This paper reports the reliability of ratings as one technic of measuring the value of the exhibits.

In the study reported, 271 professional workers comprising public health personnel, educators, students of education or medicine, and professional exhibitors; and 18,449 lay spectators rated 26 major exhibits. The professional workers designated the three "best" and the three "worst," and the lay raters indicated the three "liked most" and the three "liked least."

On only one of 26 exhibits did as many as half the professional raters agree it was among the three best exhibits shown, and even that one was rated among the worst by 11 of the 271 judges. Only 6 others were rated similarly by as many as one-fifth of the raters. The ratings indicate that the raters' criteria were not consistently applied by the 271 individuals. There was no closer agreement when the ratings were tabulated by homogeneous professional groups (public health, educator, student, or exhibitor) than when they were by the total.

On only 8 of the 26 exhibits did as many as one-fifth of the lay raters express agreement in preferences. Apparently, judgments by laymen are also unreliable as criteria to judge the effectiveness of exhibits.

Consistent with the disagreements found in the professional and lay groups were the differences between the two groups of judges. The raters who normally construct, evaluate, and use exhibits did not agree with the lay public for whom they make exhibits.

With such disagreements among the

* As an example, 2,195 laymen rated the *Maternal Health* exhibit as one of the three best, but 1,531 others rated it as one of the three worst.

ratings, it may be concluded that ratings are of little or no value in the evaluation of exhibits. Methods of evaluation which are more valid and reliable and depend more upon the behavior of the spectator must replace ratings, if the true value of health exhibits as educational media is to be appraised.*

Subsequent papers will describe the use of such technics of evaluation as: (1) timing the length of stay at an exhibit and checking it with the time required to read the legends, (2) checking the spectators' statements of the objectives of the exhibit, from viewing it, with the objectives as stated by the

exhibitor, and (3) testing information and attitudes of spectators as they have been affected by exhibits. It is hoped that similar objective studies may be undertaken of the effectiveness of pamphlets, posters, movies, and other accepted health educational technics. Only thus can we assemble the evidence necessary to justify increases in budgets for popular health education.

REFERENCES

1. Thorndike, E. L. Handwriting. *Teachers College Record*, 11, Mar., 1910.
2. Thorndike, E. L. The Measurement of Achievement in Drawing. *Teachers College Record*, 14, Nov., 1913.
3. Cohen, Joseph. The Use of Objective Criteria in the Measure of Drawing Ability. *Ped. Sem.*, 27, June, 1920.
4. Hillegas, M. B. Scale for the Measurement of Quality in English Composition by Young People. *Teachers College Record*, 13, Sept., 1912.

* A few suggestions on exhibit technic arising from studies of spectator reaction have already been reported in this *Journal* for March, 1941.