DOCUMENT RESUME

ED 092 107 40 IR 000 666

AUTHOR Bond, Jack H.
TITLE Effects of Mixed and Unmixed Criterion Questions. Observational Report No. 7226.
INSTITUTION Syracuse City School District, N.Y.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C. Media Services and Captioned Films Branch.
PUB DATE Jun 73
CONTRACT OEC-9-423617-4357(616)
NOTE 9p.; Supporting document for the Computer Based Project for the Evaluation of Media for the Handicapped

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Educable Mentally Handicapped; *Exceptional Child Research; Intermediate Grades; Measurement Techniques; *Media Research; Primary Grades; *Research Methodology; Sampling; Student Testing; *Testing
IDENTIFIERS *Computer Based Project Evaluation Media Handicap

ABSTRACT

Incorporated in the media evaluation model developed by the Computer Based Project (Syracuse, N.Y.) is a procedure whereby independent samples are used to obtain baseline responses to cognitive question items. Items from two films are intermixed and presented after showing only one of the films; thus, the sample audience responds to questions from both the unseen film and the film which was seen. This study sought to determine whether the procedure affects the responses of the samples. Statistical analyses are performed on the data gathered from educable mentally retarded (EMR) and normal samples receiving mixed questions sets. This is in contrast to EMR and normal samples receiving questions only on the film seen. The results indicate that there seems to be an effect of mixing items which results in lower scores for EMR children on the criterion items for the film seen than when only items pertaining to the film were asked. One explanation for this finding is that greater respondent frustration can be expected from the mixing of relevant (film seen) and irrelevant (film not seen) items, resulting in a lower number of correct scores. (Author/WCM)

Observational Report #7226
Jack H. Bond
June 1973

EFFECTS OF MIXED AND UNMIXED CRITERION QUESTIONS

ABSTRACT

Incorporated in the media evaluation model developed by the Computer Based Project (Syracuse, N. Y.) is a procedure whereby independent samples are used to obtain baseline responses to cognitive question items. Items from two films are intermixed and presented after showing only one of the films; thus, the sample audience responds to questions from both the unseen film and the film which was seen. This study sought to determine whether the procedure effects the responses of the samples. Statistical analyses were performed on the data gathered from educable mentally retarded (EMR) and normal samples receiving mixed question sets and from EMR and normal samples receiving questions only on the film seen. The results indicate that there seems to be an effect of mixing items which results in lower scores for EMR children on the criterion items for the film seen than when only items pertaining to the film were asked. One explanation for this finding is that greater respondent frustration can be expected from the mixing of relevant (film seen) and irrelevant (film not seen) items, resulting in a lower number of correct scores.

SPECIAL REPORT No. 7226
COMPUTER-BASED PROJECT for the EVALUATION of MEDIA for the HANDICAPPED

Title: EFFECTS OF MIXED AND UNMIXED CRITERION QUESTIONS

BACKGROUND

The Computer Based Project for the Evaluation of Media for the Handicapped, based on contract #OEC-9-423617-4357 (616) between the Syracuse (N.Y.) City School District and the Media Services and Captioned Films Branch, Bureau of Education for the Handicapped (United States Office of Education) for the five year period July 1, 1969 through June 30, 1974. The major goal is to improve the instruction of handicapped children through the development and use of an evaluation system to measure the instructional effectiveness of films and other materials with educable mentally handicapped (EMH) children, in-service training and media support for special teachers, and studies related to the evaluation process and the populations used.

The Project has concentrated on the 600 films and 200 filmstrips from the Media Services and Captioned Films (BEH - USOE) depository; however, specific packages from Project LIFE, various elementary math curricula, and selected programs from Children's TV Workshop have also been evaluated. The evaluation model used requires that: 1) objectives of materials be specified and written; 2) instruments be constructed to test and measure effectiveness; and, 3) children be the major sources of evaluation information. A number of instruments and methodologies are employed in the gathering of cognitive and affective data from 900 EMH children and 80 special teachers to make the effectiveness decisions. Over half of the EMH population can neither read or write; therefore, a unique Student Response System (SRS) is employed, consisting of a twenty station G.E.-1000 SRS which can be operated in a group or individual recording mode and is connected to a remote computer system. The computer capabilities consist of remote telephone connections to the Rome (N.Y.) Air Development Command, the Honeywell time-shared network, and the Schenectady (N.Y.) G E Research and Development Center; and batch mode capabilities of the Syracuse City Schools, Syracuse University, and various commercial sources.

In-service and media support activities provide on-the-job training for teachers, teacher aides, equipment, and materials to the special teachers in the city schools. The research activities have centered around investigations and special problems related to the development of the evaluation model. The four major areas considered are: 1) testing effects, 2) captioning effects, 3) special student characteristics; and, 4) evaluation procedures validation.

Documentation of the major activities appear in the five annual reports and the 600 evaluations prepared on materials used. Staff members were encouraged to prepare special reports and the attached paper is one of these. The opinions expressed in this publication do not necessarily reflect the position or policy of the Computer Based Project, the United States Office of Education, or the Syracuse City School District, and no official endorsement by any of the agencies should be inferred.

Observational Report #7226
Jack H. Bond
June 1973

EFFECTS OF MIXED AND UNMIXED CRITERION QUESTIONS

At the Computer Based Project a procedure of using independent samples to obtain base line responses to cognitive criterion items has replaced one which used pretest/posttest data on the same sample for each film showing (CBP, 1972). The new procedure was instituted to shorten the time required to administer question items; it requires that items from two films be intermixed, and presented after showing only one of the films. Thus the sample audience responds to questions from both the unseen film and the seen film after being exposed to only one film to which the items apply.

The question has been raised: "What effects have these unrelated items on those items that are related to the film seen?" The specific hypothesis is there are no significant differences (P = .05) between the correct response of groups receiving only the items for the film seen and those receiving mixed item sets.

METHOD

It was observed that through an administrative error several sample audiences consisting of EMR and normal children in primary and intermediate "intact" classrooms were shown the film #1426 "Sniffy Escapes Poisoning," and asked only those questions pertaining to the film seen from the scheduled mixed question set. Other samples from the same population were shown the film and asked all of the items in the mixed question set, containing both items for #1426 (the film seen) and for

a film not seen. The strata for items and for two academic groups were represented in the available samples: educationally mentally retarded (Group I, mixed items, and Group II, only #1426 items, and regular (normal CA 8-12) children Group III, mixed items). Both

TABLE I

MIXED AND UNMIXED ITEMS PERCENT CORRECT AND RANKS FOR EMH AND NORMAL SAMPLES ON FILM #1426

GROUP	ITEMS: PERCENT CORRECT 1	2	3	4	5	6	7	8	9	10	MEAN	SUM RANK
EMH (GROUP I) (1) Mixed	68	55	55	68	64	45	73	59	50	50	52	
A. Ranks I & II	15.5	7.5	7.5	15.5	13	3	17	12	5	5		101
B. Ranks I & III	13.5	9.5	9.5	13.5	12	3	15	11	5.5	5.5		98
C. Ranks I, II, & III	20.5	12.5	12.5	20.5	18	5	22	17	8	8		144
EMH (GROUP II) (2) Unmixed	58	75	75	58	33	75	58	17	67	50	56	
D. Ranks I & II	10	19	19	10	2	19	10	1	14	5		107
E. Ranks II & III	10	14	14	10	2	14	10	1	12	6		93
F. Ranks I, II, & III	15	24	24	15	2	24	15	1	19	8		147
MIXED (GROUP III) (3) Regular	36	36	50	54	50	100	86	100	86	93	69	
G. Ranks I & III	1.5	1.5	5.5	8	5.5	19.5	16.5	19.5	16.5	18		112
H. Ranks II & III	1.5	1.5	6	8	6	19.5	16.5	19.5	16.5	18		117
I. Ranks I, II, & III	3.5	3.5	8	11	8	29.5	26.5	29.5	26.5	28		179

Group I and III received the items in mixed order, and Group II received an unmixed set of items.

The percent correct response for those items for #1426 in these two situations is summarized in Table I by percent of respondents giving a correct response to the item and the rank of the score when combining the item per correct scores to rank order from lowest to highest in four comparisons -- I with II, II with III, III with I, and all three. For example, the "68" for item 1, Group I, in Table I is 15th when ranked with Group II scores; 13th when ranked with Group III scores; and 20th when all three groups are ranked.

The above item percentage scores indicate a nominal order, continuous distribution for each sample group. The effects of mixing items and differences between EMR and normal populations were tested using a 3 group Kruskal-Wallis one-way analysis of variance on the ranked item percent correct scores (Seigel, p. 184-193, 1956). The purpose was to test the null hypothesis that (H_o) there is no difference between the average correct percent score of the three samples, and an alternative (H_1) that the 3 samples are different in the average percent correct scores. The value of the statistic H was computed from the sum of the ranks for each sample using the values from lines C, F, and I, of Table I above in the H formula (Seigel, p. 187). The correction factor of .993 for seven pairs of tied values present in the data was also computed. The resulting H = 3.00 was appropriate for p = .30 which is greater than the rejection value of p = .05 (Table C, Seigel, p. 249). Thus the null hypothesis cannot be rejected

and indicates that the sample groups are not significantly different in the average percent correct score.

$$H = \frac{12}{30\,(30+1)}\left(\frac{[144]2}{10} + \frac{[147]2}{10} + \frac{[179]2}{10}\right) - 3(30+1) = 2.9819$$

$$\text{Corrected } H = \frac{H}{1 - \left(\frac{5[2^3 - 2] + [5^3 - 5] + 2\,[3^3 - 3]}{([30]^3 - 30)}\right)} = 3.0012$$

The finding suggests that data from any one of these samples could be used to make item inclusion decisions, or estimates of population responses, and that the responses of the groups are not significantly different from one another. It could be concluded that the effects of mixing items had no significant effect upon the resulting scores for items for the film seen.

The differences in group means from Table I indicate that the mixed sets for EMR children have the lowest mean score, and that normal groups have the highest mean score. This finding is expected. It was assumed that a greater respondent frustration would be expected from the mixture of relevant (about film seen) and 12 irrelevant (unseen film) items resulting in a lower correct score. A number of studies (CBP 711, 718, 725, Micro-Exp.) have suggested that normal groups do better than EMH groups on academic tasks.

The group mean differences were submitted to a Mann-Whitney U test (the non-parametric substitution for a t-test of means when data is ordinal), to determine if the differences were significant. A tabled

value for U = 27, p = .05, for groups of 10 is required to reject the null hypothesis that means are the same (Table K, Seigel, 1956, p. 277). An obtained value of between group comparisons was calculated from the ranks in Table I resulting in the following values: $U_{12} = 46$; $U_{23} = 38$; $U_{13} = 43$. All of these values exceed the tabled value, and suggest the mean of each group is different from the other two groups. Adding this finding to the K-W finding of the samples from the same population, indicates the samples did perform differently.

If one uses a criterion of 50% correct response on the item as acceptable (CBP, 1972) for inclusion in the evaluation instrument, it can be noted that one item (#6) would be rejected if the mixed sample were used (Group I, Table I); two items (#5, #8) would be rejected by the unmixed sample (Group II, Table I); and two items (#1, #2) by the regular Group III. The inconsistency of the specific items that might be rejected seems to be due to unexplained sample variations; however, each would reject at least one item. The percentages across all three groups for item #5 tend to be somewhat in the lower range of each set (4th in Group I, 9th in Group II, and 6th in Group III); however, the responses for #8 and #1 do not seem to fit any pattern; in fact, a reversal is evident. Item #1 is high for Group I and low for Group III; whereas, item #8 is high for Group III and low for Group II.

This conclusion is further statistically verified by computing Spearman rank order correlations using the items by groups indicates the variability of item scores between groups. If each group received the same proportion of the group score from each of the items the correlations should be near 1.0. The results are shown in Table II below.

TABLE II

SPEARMAN RANK ORDER CORRELATION

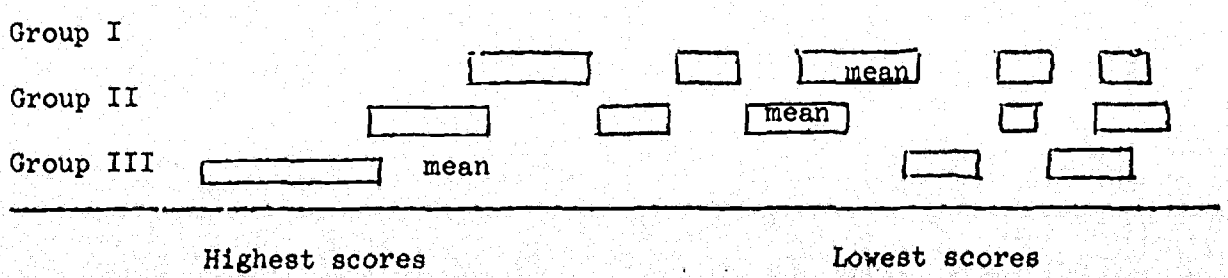
	I	II	III
Group I		-.40	-.20
Group II			-.40

Note the correlations are all negative, suggesting the inverse ranking relationship between the groups noted in the paragraph above. Also the contributions of individual items to the total score of each sample group are not proportional and equally distributed.

A graphic presentation of the dispersion of scores is shown below, suggesting the reason for the apparent discrepancies between the K-W and Mann-Whitney findings as being one of dispersion of the scores.

FIGURE I

DISPERSION OF SCORES FOR 3 GROUPS



IMPLICATIONS

There seems to be an effect of mixing items which results in lower scores for EMR children on the criterion items for the film seen than when only items pertaining to the film are asked. In a system where future decisions are made at the 50% level, the item decisions tend to be on the conservative side thus resulting in the discarding of some good items, and probably results in lower achievement from mixed item sets than would be indicated when only the items for the film seen are used.

REFERENCES

Computer Based Project Notes, CBP for Evaluation of Media for the Handicapped. Special Education Department, City School District, Syracuse, New York 1972.

Computer Based Project Research Report #711: CA and achievement factors in special students response to media. Computer Based Project for the Evaluation of Media for the Handicapped, City School District, Syracuse, New York: January, 1971.

Computer Based Project Research Report #718: Sample population sub-group comparisons. Computer Based Project for Evaluation of Media for the Handicapped, City School District, Syracuse, New York: June, 1971.

Computer Based Project Research Report #725: Functional basic word list vocabulary study. Computer Based Project for the Evaluation of Media for the Handicapped, City School District, Syracuse, New York: March, 1972.

Computer Based Project Research Report #704: Micro-experiments. Computer Based Project for the Evaluation of Media for the Handicapped. City School District, Syracuse. New York: February, 1971.

Seigel, Sidney: Non-parametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company, Inc., 1956.