DOCUMENT RESUME

ED 029 701 PS 001 822

By-Larkins, A. Guy; Shaver, James P.

Comparison of Yes-No, Matched-Pairs, and All-No Scoring of a First-Grade Economics Achievement Test.

Pub Date 12 Apr 68

Note-12p.; Paper presented to the Education Section of the Utah Academy of Sciences, Arts, and Letters, Provo, Utah, April 12, 1968

EDRS Price MF-$0.25 HC-$0.70

Descriptors-*Achievement Tests, Comparative Analysis, *Primary Grades, *Test Construction, Testing, *Testing Problems, Test Reliability, Test Validity

Identifiers-All No Tests, Matched Pairs Scoring, Yes No Tests

Developing practical achievement tests for use at the primary-grade level is a difficult task. Some problems encountered appear to be resolved by using verbally administered yes-no tests. But such tests are criticized as having a low reliability because they offer only two choices. Two modifications of the yes-no test have been proposed to increase reliability. One is the "matched-pairs" technique, in which every "yes" item has a matching item to be answered "no". Both items must be answered correctly for either to be counted. The second technique of the all-no test, an attempt to counter the children's proclivity to answer "yes" even when the answer is not known. Some 200 first grade children were administered an economics test, in which all three techniques were used. The test scores indicated that the all-no test had the greatest reliability, but it was less valid than the matched-pairs test. Thus, the matched-pairs test would be the best way to construct the yes-no type of achievement tests. Another article by the same authors (see PS 001 819) also deals with the subject. (WD)

COMPARISON OF YES-NO, MATCHED-PAIRS, AND ALL-NO SCORING OF A FIRST-GRADE ECONOMICS ACHIEVEMENT TEST*

by

A. Guy Larkins

and

James P. Shaver
Utah State University

---

*Paper presented to the Education Section of the Utah Academy of Sciences, Arts, and Letters. Provo, Utah, April 12, 1968.

PROBLEM

Developing practical achievement tests for use at the primary-grade level is a difficult task for teachers or researchers. Written multiple-choice tests require reading ability on the part of the child. Multiple-choice picture tests require too much time to construct. Interviews require too much time to administer, and they are difficult to standardize.

These problems could be surmounted by using verbally administered YES-NO tests.[1] Such tests, however, have low reliability because they offer the student only two options on each item. They also are difficult to interpret because they are sensitive to acquiescence response set[2] (Shaver and Larkins, 1966; Larkins and Shaver, 1967).

Larkins and Shaver (1967) reported an attempt to correct for response set and increase the reliability of the YES-NO test. They produced a YES-NO Primary Grades Economics Test (called PET-1) which was written with reversals and scored using matched-pairs. That is, for every item for which the correct response was YES (YES items), a matching item was written for which the correct response was NO (NO items); and the student was required to correctly respond to both the YES item and the NO item in a matched-pair before receiving credit for either item.

It was expected that this scoring technique would increase the reliability of the YES-NO test by decreasing the probability of a correct chance response from one-in-two to one-in-four. It was also expected that this technique would correct for response set. If the student tended to guess

---

[1]After the teacher reads an item, the students mark either YES or NO on their printed answer sheets.

[2]Acquiescence response set is the tendency to respond YES when not responding from knowledge. In this paper, the tendency to respond NO is called dissent response set.

YES, he would miss the NO items; if he tended to guess NO, he would miss the YES items.

The PET-1 test was administered to experimental and control groups of first-grade children--students who had been instructed in economic concepts, and students who had not. The test was then scored in both the ordinary manner and using the Matched-Pairs technique. In both the experimental group and in the control group, Matched-Pairs scores were more reliable than ordinary YES-NO scores.[3]

Although Matched-Pairs scoring was successful in increasing reliability, the corrected Matched-Pairs split-half reliability coefficient of .60 was barely adequate for comparing group means and fell short of the .90 usually considered desirable for discriminating between individual scores. Even though reliability might have been increased by increasing the length of the test, other approaches to improving the YES-NO test were also investigated Cronbach (1942) suggested that the reliability of the YES-NO tests could be increased by writing tests containing only NO items. Since most people tend to acquiesce rather than dissent, a NO response would more frequently be made from knowledge than would a YES response. However, NO-item-only tests (All-NO tests) probably favor the dissenter--a person who tends to respond NO when not responding from knowledge would obtain a spuriously high score on such an achievement test.

No comparison of the reliability and validity of YES-NO, Matched-Pairs,

---

[3] The corrected split-half reliability coefficient for scores based on thirty YES-NO items was .35 in the experimental group. In the same group, the corrected coefficient for scores based on fifteen matched-pairs of items was .60. Coefficients for YES-NO and Matched-Pairs scores in the control group were .14 and .46. Split-half reliability coefficients as high as .85 have been obtained with Matched-Pairs scoring on longer tests.

and All-NO test scores was available in the literature, although this information would be important to researchers and teachers when selecting test formats.

OBJECTIVES

The objectives of this study were to:

1. Devise a first-grade economics achievement test or tests which would yield YES-NO, Matched-Pairs, and All-NO scores, and

2. Determine whether these scores differed in reliability and validity.

EXPECTATIONS

Based on previous experience with YES-NO and Matched-Pairs tests, Cronbach's advice concerning YES-NO and All-NO tests, and limited experience with one All-NO test, it was expected that in reliability the tests would rank: All-NO, Matched-Pairs, and YES-NO, with All-NO the highest.

Expectations concerning validity were based on the a priori argument that All-NO scores for highly acquiescent students would be lower than for less acquiescent but equally knowledgeable students. That is, that All-NO scores would confound response set and knowledge. It was expected that if this confounding were serious, certain predictions based on the construct of knowledge[4] would not be confirmed with the All-NO test. Those predictions were:

1. PET-1 achievement scores for knowledgeable groups will be more reliable than for ignorant groups.

---

[4]These predictions were based on the assumption that students who received an experimental treatment would be knowledgeable compared to students who did not receive the experimental treatment. As reported elsewhere, this assumption appears to have been sound (Larkins, 1968).

2. PET-1 achievement scores for knowledgeable groups will be more variable than for ignorant groups.

3. PET-1 means for knowledgeable groups will be larger at the .01 level of significance than for ignorant groups.

PROCEDURE

Two Primary Economics Tests were written based on Our Working World: Families at Work (Senesh, 1963). The first was a 74 item, All-NO test. The second was a similar YES-NO test, written with reversals so that it could be scored using the Matched-Pairs technique. These tests were administered, as part of a larger study (Larkins, 1968), to two experimental groups and one control group. Students were selected as classroom units without randomization. Three experimental classes and three control classes were selected from two school districts in northern Utah. Three experimental classes were selected from the Elkhart, Indiana, School District. Although students were not selected randomly, Tests of General Ability (Flanagan, 1959) scores were obtained and used to correct for initial differences in mental ability when PET-1 means were compared.

Split-half reliability coefficients, corrected with the Spearman-Brown Prophecy Formula, were computed and compared for sets of YES-NO, Matched-Pairs, and All-NO scores. This was done to test the expectation concerning reliability, and to test the first prediction related to validity. This was the only part of the analysis in which the 74 item, All-NO test was used. When means, standard deviations, F-tests, and t-tests were computed to test the second and third predictions related to validity, it was necessary to hold constant all other factors except test form. This was accomplished by obtaining YES-NO, Matched-Pairs, and All-NO scores from a single administra-

tion of the YES-NO test. The All-NO scores for this part of the analysis were obtained, therefore, by using only the NO items on the YES-NO test.

FINDINGS

Findings presented in Table 1 support the expectation concerning reliability. As predicted, in reliability the three sets of scores ranked: All-NO, Matched-Pairs, and YES-NO.

Table 1. Split-half reliability coefficients for YES-NO, Matched-Pairs, and All-NO scores.

Group	N[a]	YES-NO 75 Items[b]	YES-NO Matched-Pairs 37 Pairs	All-NO 74 Items
1	77	.68	.85	.90
2	59	.48	.66	.89
3	77	.29	.62	.87

[a]The number of students in the group.
[b]The number of items on the test.

It was expected that certain predictions based on the construct of knowledge would not be confirmed with the All-NO test. Three groups were available with which to test these predictions. In Table 1, Group 1 is an experimental group which was taught economic concepts under optimal conditions. The teachers had both special training and previous experience in using Families at Work, and the students' mean score on the Tests of General Ability was close to one grade level above their grade at the time of testing. Group 2 is also an experimental group, but was taught economics under more nearly average conditions. The teachers had neither special training nor experience in using Families at Work, and the group's mean score on

the mental abilities test was not above grade level. Group 3 is the control group. These children received no instruction with the Families at Work program. On three sets of PET-1 scores, the groups ranked 1, 2, and 3 in knowledge of economic concepts (see Table 2).

The first prediction related to validity was that PET-1 scores for knowledgeable groups would be more reliable than for less knowledgeable groups. It can be seen in Table 1 that the reliability coefficient for the All-NO test is nearly as large for the least knowledgeable group as it is for the most knowledgeable group. This was not true of the YES-NO or Matched-Pairs tests. Because the All-NO test produced such stable reliability coefficients, its validity must be questioned. A two-option test which produces reliability coefficients which do not vary from knowledgeable to ignorant groups is probably testing something other than, or in addition to, knowledge.

The second prediction based on the construct of knowledge was that PET-1 scores for knowledgeable groups would be more variable than for less knowledgeable groups. Findings presented in Table 2 indicate that this prediction was confirmed with the YES-NO scores and the Matched-Pairs scores, but not with the All-NO scores.[5]

---

[5]The difference between standard deviations for Groups 1 and 3, checked using the variance ratio, was significant at the .05 level for the YES-NO test, and at the .01 level for the Matched-Pairs test. Since the predictions were directional, a one tailed test of significance was used.

Table 2. Means and standard deviations for YES-NO, Matched-Pairs, and All-NO scores derived from a single administration of the YES-NO test.

Group	N[a]	YES-NO[b] ½M	YES-NO[b] ½SD	YES-NO Matched-Pairs M[c]	YES-NO Matched-Pairs SD[d]	All-NO M	All-NO SD
1	77	27.75	3.76	20.46	6.58	25.16	6.31
2	59	24.15	3.38	15.14	5.41	20.05	6.58
3	46	23.30	2.97	13.65	4.78	19.15	6.46

[a]The number of students in the group.

[b]The YES-NO test is twice as long as the others. In order to make a direct comparison, its means and standard deviations were reduced by half.

[c]The mean.

[d]The standard deviation.

Two observations are of particular interest in regards to Table 2.

1. For all three groups, the YES-NO test tends to be less variable than either of the other tests--its standard deviations are smaller.[6] One explanation might be that, since students tend to be acquiescent, YES items obscure differences between ignorant and knowledgeable students. Both respond YES; one from knowledge, the other from response set. The reduction in variability among students on the YES items may reduce the standard deviation for the total test.

2. In Groups 2 and 3--the least knowledgeable groups--the standard deviations for the All-NO test are larger than for the Matched-Pairs test.[7]

---

[6]The differences between the standard deviations for the YES-NO test and each of the other two tests were significant at the .01 level in each instance. However, the computation did not take into account the correlation of the test scores which is caused by using a single group. The significance of the difference between standard deviations is, therefore, probably even higher.

[7]The difference between standard deviations was significant at the .05 level for Group 2, and at the .01 level for Group 3. However, since the formula for correlated scores was not used, the significance of the difference between standard deviations is probably even higher.

Furthermore, the standard deviations for the All-NO test are similar in all three groups, but the standard deviations for the Matched-Pairs and the YES-NO test decrease from Groups 1 to 3.

The All-NO and YES-NO standard deviations are probably spurious if taken as indicators of variability in knowledge. Scores on the YES-NO test apparently are less variable in all groups than they would be if the instrument were not measuring dissent in addition to knowledge.

The third prediction based on the construct of knowledge was the PET-1 means for knowledgeable groups would be larger at the .01 level of significance than for ignorant groups. As indicated by the standard deviations in Table 2, the variability of YES-NO and All-NO scores is affected by response set as well as knowledge. Since parametric tests of significance utilize sample variance, i.e., the standard deviation, to estimate population variance, it is possible that when acquiescence is confounded with knowledge, groups might appear to differ in knowledge when they do not, or groups might appear not to differ in knowledge when they do differ.

The significance of the difference among PET-1 means for the three groups was tested using analysis of covariance, with adjustments for initial differences in mental ability.[8] The significance of the differences between PET-1 means for pairs of groups was then tested using the t-test. Table 3 presents the t-ratios for the three PET-1 tests and the three groups.

---

[8] F-ratios comparing all three groups on the YES-NO, Matched-Pairs, and All-NO tests were 19.34, 16.03, and 9.16. $F_{.01} = 4.71$.

Table 3. T-ratios between PET-1 means adjusted for initial differences in mental ability.

Groups	Adjusted YES-NO	Adjusted Matched-Pairs	Adjusted All-NO
1 and 2	1.96	1.10	.54
2 and 3	2.53	2.81	2.32
1 and 3	4.53	4.00	3.00

$t_{.05}=1.98$ $t_{.01}=2.61$

Since Groups 1 and 2 received the experimental treatment, it was expected that their PET-1 means would not differ. Since Group 3 did not receive the experimental treatment, it was expected that its PET-1 means would differ from those for both Group 1 and Group 2. The findings in Table 3 indicate that Groups 1 and 3 differed at the .01 level of significance for all three PET-1 tests, as expected. However, results with the YES-NO test very nearly failed to confirm the expectation that the PET-1 means for Groups 1 and 2 would not differ--the t-ratio is nearly significant at the .05 level. And, both the YES-NO test and the All-NO test failed to confirm the prediction that PET-1 means for Groups 2 and 3 would differ at the .01 level of significance. In other words, the Matched-Pairs test was the only one to uniformly confirm the prediction that PET-1 means for knowledgeable groups of students would significantly differ from PET-1 means for less knowledgeable groups of students.

SUMMARY

Larkins and Shaver (1967) reported that the validity and reliability of the YES-NO type test could be improved by writing items with reversals and scoring the test with matched-pairs. Cronbach (1942) recommended that the

reliability of the YES-NO type test could be improved by including only NO items. However, on a priori grounds it appeared that the All-NO test would produce spuriously high achievement scores for students who were not acquiescent. Furthermore, no direct comparison had been made between Matched-Pairs tests and All-NO tests for validity and reliability.

In the present study, YES-NO, Matched-Pairs, and All-NO scores were compared. It was concluded that the All-NO test had greater reliability than the YES-NO or Matched-Pairs tests. It was also concluded, however, that the All-NO test was less valid than the Matched-Pairs test. This conclusion was based on the lack of confirmation, with the All-NO test, of three predictions based on the construct of knowledge. One of these predictions was also not confirmed with the YES-NO test.

Based on these results, researchers and primary-grade teachers would be well-advised to use Matched-Pairs scoring when writing YES-NO type achievement tests.[9]

---

[9]Research utilizing several Matched-Pairs tests (Larkins, 1968) has indicated that to obtain adequate reliability for establishing grades for individual primary grade students, at least 60 pairs of items will frequently be needed.

LITERATURE CITED

Cronbach, Lee J. 1942. Studies of acquiescence as a factor in the true-false test. Journal of Educational Psychology 33(6):401-415.

Flanagan, John C. 1959. Tests of general ability: Form A grades K-2. Science Research Associates, Inc., Chicago, Illinois.

Larkins, A. Guy, and James P. Shaver. 1967. Matched-pairs scoring technique used on a first-grade YES-NO type economics achievement test. Utah Academy of Sciences, Arts, and Letters: Proceedings 44(1): 229-242.

Larkins, A. Guy. 1968. Assessing achievement on a first-grade economics course of study. EdD dissertation. Utah State University. 202 p. University Microfilms, Ann Arbor, Michigan.

Senesh, Lawrence. 1963. Our working world: Families at work, resource unit. Science Research Associates, Inc., Chicago, Illinois. 198 p.

Shaver, James P., and A. Guy Larkins. 1966. SRA economics materials in grades one and two: Evaluation report to Salt Lake City School District. Bureau of Educational Research, Utah State University. (Mimeographed)