DOCUMENT RESUME

ED 358 136 TM 019 921

AUTHOR	Huntley, Renee M.; Welch, Catherine J.
TITLE	Numerical Answer Options: Logical or Random Order?
PUB DATE	Apr 93
NOTE	8p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE	Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Ability; College Entrance Examinations; Comparative Testing; *Distractors (Tests); High Achievement; Higher Education; Item Analysis; Low Achievement; *Mathematics Tests; *Multiple Choice Tests; *Standardized Tests; *Test Construction; Test Format; Testing Problems; Test Items; Word Problems (Mathematics)
IDENTIFIERS	*ACT Assessment; Item Position (Tests); *Numerical Order; Randomization; Speededness (Tests); Test Developers

ABSTRACT

Writers of mathematics test items, especially those who write for standardized tests, are often advised to arrange the answer options in logical order, usually ascending or descending numerical order. In this study, 32 mathematics items were selected for inclusion in four experimental pretest units, each consisting of 16 items. Two versions presented options in logically descending or ascending order, and two presented them randomly. Each test was administered as part of the American College Test (ACT) pretest procedures to approximately 300 examinees. A review of the speededness rate indicates that 99 percent of the examinees completed the test, leading to the conclusion that not only the difficulty, but the speededness was unaffected by the random ordering of the numerical options. Regarding the higher biserials in performance on the randomly ordered options, the higher-ability examinees (as determined by their performance on the ACT Assessment Mathematics Test) scored proportionately better than did the lower-ability group. It is evident that, while random ordering of numerical options is not an obstacle for upper-ability examinees, it may be one for lower-ability examinees. Given the difference in biserials, test writers might do well to avoid introducing confusion or an irrelevant source of error by continuing to arrange options in some logical sequence. Two tables present study findings. (SLD)

Numerical Answer Options: Logical or Random Order?

Renee M. Huntley
Catherine J. Welch

American College Testing

Paper presented to the annual meeting of the
American Educational Research Association
Atlanta, April 1993

Numerical Answer Options: Logical
or Random Order?

Objectives

Writers of mathematics test items, especially those who write test items for standardized testing companies, are often admonished to arrange answer options in logical order. They are urged to order options, when possible or meaningful, in some logical sequence, usually ascending or descending numerical order. These editorial and format concerns are illustrated in the following item:

> How much larger than the sum of -3, -7, and 5 is the product of these same integers?
>
> A. -110
> B. -100
> C. 90
> D. 100
> E. 110

Some introductory measurement texts categorically state that "Numerical options...should always be placed in ranked sequence, either ascending or descending" (Sax, 1974); others merely observe the issue in the context of the arrangement of options "as simply as possible. If numbers, they should be in ascending or descending order; if single words, alphabetically; if phrases, order of length" (Mehrens & Lehmann, 1984). At least one text hypothesizes that this practice of forming a "quantitative scale...may avoid some confusion on the part of the examinee and eliminate an irrelevant source of error" (Ebel and Frisbie, 1991). None of these text books, however, seem to provide any empirical base for the stricture.

In contrast, there is a relatively large body of research, some of it quite recent, concerning the issue of placement of keyed responses and ordering of answer options (Cizek, 1991; Cronbach, 1950; Fagley, 1987; Friel and Johnstone, 1979; Huntley and Welch, 1988; Jessel and Sullins, 1975; Marcus, 1963). Within this body of research, however, there are mixed results regarding performance effects on difficulty and discrimination. Furthermore, a literature search did not uncover any research that specifically addressed the performance effects of randomly ordered numerical responses. This issue is therefore the subject of inquiry of the present study, which attempts to provide an empirical basis for the conventional wisdom. Many measurement admonitions have been traditionally expounded because they have seemed so intuitively logical and right. Some, however, upon empirical examination, have proven to be unfounded.

This study provides yet another instance for examination. One could speculate that ordered numerical options would be of particular importance, for example, in a timed test, where the search for the numerically appropriate option would be impeded by a random order. (The parallel, obviously, would be the search for a word in an unalphabetized dictionary.) The results of this study, however, do not seem entirely to support these assumptions.

Method

Thirty-two mathematics items were selected for inclusion in four experimental pretest units, each consisting of sixteen items. Version A items presented distractors in logically ascending or descending order; Version B presented distractors for the Version A items in random order. Version C items similarly presented distractors in logically ascending or descending order; Version D presented the randomly ordered distractors for the Version C items.

Each item was administered as part of the ACT Assessment pretest procedures to approximately 300 examinees in randomly equivalent samples. Only one pretest unit was administered to any one examinee, and no examinee was given both versions of the same item. All four units were administered in the same pretest administration year.

Differences in performance (p-values, biserials, and omits) between the two versions of the items were examined. It should be noted that biserials were derived by using the total 40-item test score on the ACT Assessment Mathematics Test, which was administered at the same time. Performance on the ACT Assessment Mathematics Test was also used to define the following examinee ability levels: the upper 27% of examinees constituted the high-ability level; the middle 45%, the middle-ability level; and the lower 27%, the low-ability. Speededness was defined as the percentage of examinees who completed the last five items in all experimental units.

Results

Tables 1 and 2 provide the item statistics for the 32 items included in this study. The original order refers to the performance of the items when the distractors had been ordered to form a quantitative scale (that is, in either ascending or descending order). The random order refers to the performance of the items when the distractors had been randomly ordered.

Review of Tables 1 and 2 indicates that there was no significant difference in p-values when the order of the distractors changed. For the items in Units A and B, the mean p-value decreased slightly from 47.56 to 46.63. The opposite was true for the items in Units C and D, where the mean p-value increased slightly from 36.19 to 36.82. The largest change was observed in the biserial correlations, which increased substantially from the original order to the random order. For Units C and D, the mean biserial increased from 34.25 to 43.31. For Units A and B, however, the increase was less (45.81 to 47.81), but still positive.

Discussion

A review of the speededness rate indicated that 99% of the examinees completed the test. We could conclude, therefore, that not only the difficulty (p-values) of the items, but the speededness was unaffected by the random ordering of the numerical options.

Regarding the higher biserials in performance on the randomly ordered options, the higher-ability examinees (as determined by their performance on the ACT Assessment Mathematics Test) scored proportionately better than the lower-ability group. Consequently, it is clear that while random ordering of numerical options is not an obstacle for upper-ability examinees, it may be for lower-ability ones.

What should a test constructor do? Given the difference in biserials, test writers might well opt to avoid introducing "confusion" or "an irrelevant source of error" (Ebel and Frisbie, 1991) by continuing to arrange numerical options in some logical sequence.

Bibliography

Cizek, G.C. The effect of altering the position of options in a multiple-choice examination. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1991.

Cronbach, L.J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.

Ebel, R.L. & Frisbie, D.A. (1991). Essentials of Educational Measurement, Fifth Edition. Prentice Hall, 174.

Fagley, N.S. (1987). Positional response bias in multiple-choice tests of learning: its relation to testwiseness and guessing strategy. Journal of Education Psychology, 79, 95-97.

Friel, S., & Johnstone, A. H. (1979). Does the position matter? Education in Chemistry, 16, 175.

Huntley, R.M. & Welch, C. The effect of answer location on item difficulty and discrimination in language-usage tests. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, April 1988.

Jessell, J.C. & Sullins, W.L. (1975). The effect of keyed response sequencing of multiple-choice items on performance and reliability. Journal of Educational Measurement, 12, 45-48.

Marcus, A. (1963). The effect of correct response location on the difficulty level of multiple-choice questions. Journal of Applied Psychology, 47, 48-51.

Mehrens, W.A. & Lehmann, I.J. (1984). Measurement and Evaluation in Education and Psychology, Third Edition. Holt, Rinehart and Winston, 174.

Sax, G. (1974). Principles of Educational Measurement and Evaluation. Wadsworth Publishing Company, 93.

Table 1
Comparison of P-values and Biserials for Units A and B

Item #	P-value		Biserial	
	Original (A)	Random (B)	Original (A)	Random (B)
1	65	56	59	72
2	58	50	62	62
3	82	84	49	41
4	31	27	35	33
5	38	36	39	57
6	31	33	37	38
7	25	27	34	25
8	28	33	59	27
9	53	52	44	57
10	41	37	50	67
11	53	50	64	65
12	46	51	44	37
13	32	33	48	56
14	49	51	58	65
15	87	87	21	27
16	42	39	30	36
Mean	47.56	46.63	45.81	47.81
SD	17.79	17.31	12.18	15.78

Table 2
Comparison of P-values and Biserials for Units C and D

Item #	P-value		Biserial	
	Original (C)	Random (D)	Original (C)	Random (D)
1	33	42	43	63
2	89	88	39	34
3	38	43	39	36
4	16	15	18	41
5	20	27	26	18
6	41	39	28	41
7	13	9	32	47
8	41	45	66	73
9	31	24	3	2
10	35	26	28	42
11	39	40	39	62
12	43	43	42	51
13	37	40	59	59
14	44	44	29	50
15	52	56	40	59
16	7	8	17	15
Mean	36.19	36.82	34.25	43.31
SD	18.20	18.86	14.95	18.59