

1958

The effect of correct response location on the item difficulty level of multiple-choice questions.

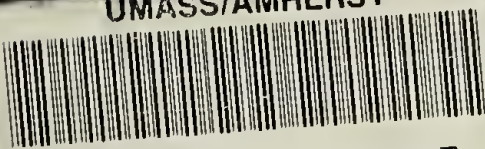
Arthur. Marcus
University of Massachusetts Amherst

Follow this and additional works at: <https://scholarworks.umass.edu/theses>

Marcus, Arthur., "The effect of correct response location on the item difficulty level of multiple-choice questions." (1958). *Masters Theses 1911 - February 2014*. 1766.
<https://doi.org/10.7275/6871312>

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

UMASS/AMHERST



312066013052867

THE EFFECT OF CORRECT RESPONSE LOCATION
ON THE ITEM DIFFICULTY LEVEL OF
MULTIPLE-CHOICE QUESTIONS



MARCUS - 1958

LD
3234
M268
1958
M322

THE EFFECT OF CORRECT RESPONSE LOCATION
ON THE ITEM DIFFICULTY LEVEL OF
MULTIPLE-CHOICE QUESTIONS

Digitized by the Internet Archive
in 2014

Thesis Submitted in Partial Fulfillment of the
Requirements for the Master of Science Degree
in Psychology at the
University of Massachusetts
1958

Table of Contents

	<u>Page</u>
Introduction.....	1
Background literature.....	2
Statement of the problem.....	6
Method.....	8
Subjects.....	8
Stimulus materials.....	8
Procedure.....	9
Results.....	12
Discussion.....	23
Summary.....	27
References.....	29
Appendix.....	30
Instructions administered to groups.....	30
Randomization for choice patterns.....	31
Acknowledgments.....	32

Introduction

It has been well established by Cronbach (4,5) that a person's performance on an objective test is affected by the variety of test-taking tendencies with which he approaches the test. These test-taking tendencies he refers to as response sets. Among these sets are included the tendency to work for speed rather than accuracy, to guess when uncertain of an answer, to favor a particular response when certain fixed alternatives are offered, and the tendency to favor certain response positions to the neglect of others.

It has been shown (4,5,6,7) that both the manner in which the test items are presented and the nature of the instructions may serve to influence the subject to respond in a particular way. In such instances a change in the item form or a modification in instructions is found to alter the subject's responses. The response set, then, can be seen as a function of the test being dealt with and in fact may be totally unrelated to the ability being measured. When certain aspects of the procedure are not made clear in test instructions, response sets may differ from one person to another, thereby yielding a source of variation in the testing conditions. These personal ways of responding to test items on a specific test form contribute to invalidity. Thus response sets can be seen as weakening a test by introducing extraneous factors which were not intended to be part of its

original content but which lower its validity.

Background literature.--One of the earliest studies to be conducted on the influence of positional preference to questions in two-response types of tests was done by Mathews (6). He found that on questions which involved the indication of a preference, a tendency occurred for pupils to mark a given response more often when it was printed above an alternative response than when it was printed below it. On information questions requiring a yes or no answer there was a tendency to mark a given response more often when it was printed to the left, than when it was printed to the right. In a subsequent study, it was shown by Mathews (7) that, on a fifty item questionnaire printed in two forms and devised to discover the attitudes of pupils toward various study habits, the position of the five possible responses influenced the way in which an individual would answer items on the blank. This was indicated by the fact that a given response was chosen from .1 to 7.2 percent more often when in one position than when it appeared in another position. Position one was responded to more often than position two, and position four was responded to more often than position five.

That the difference in response sets may possibly reflect personality characteristics rather than ability is indicated by the work of Berg and others (1,2). These studies point out that extreme position response sets may

serve as possible measures of personality and group characteristics, which are stable in test-retest situations. In another study on the position factor, Rapaport and Serg (10) find that in an imaginary questionnaire where subjects were asked to check off one of four options when no actual question was stated, over sixty percent of those responding chose the third position.

Most of the above mentioned tendencies to select or neglect certain response positions have had to do with an individual's willingness to choose from among such unstructured responses as "like very much" and "dislike very much" and only serve to indicate that responses to such multiple choice questions may be an individual expression of set. However, very little work has been done to date with the conventional type of multiple choice tests dealing in a knowledge of subject matter or with scholastic aptitude. Cronbach (5) concludes that the objective type of multiple choice form is relatively free from response sets. He demonstrated this by analyzing thousands of test papers from the Hennon-Nelson Test of Mental Ability, Form A, for Grade 3-5. A bias score was obtained for each individual and its reliability was determined. No statistical evidence was found indicating a set to prefer any particular position. A similar analysis was made using a modified version of the Ohio State Psychological Examination and again no significant evidence of bias was revealed. Some contradictory

evidence exists which indicates that position preferences may occur. Writers such as McNamara and Weitzman (8) generalize from their research that a tendency exists in test-takers to select certain positions in a serial listing of choices in preference to others. Their findings suggest that the difficulty level of a multiple choice test item is influenced by the position in which the correct choice has been placed. They define difficulty as the percentage of subjects selecting the correct choice when it appears in each of the several positions. These investigators find that for five choice items, those items having right answers in the fourth position are the most difficult, those with right answers in the second and third position are the easiest, and the first and fifth positions are of equal moderate difficulty. Their results for four choice items show that the third position is most difficult, and that difficulty increases from the first through the third position, and decreases with the fourth. This finding is interpreted as agreeing with the results for five choice items in that the next to the last position is always found to be most difficult. McNamara and Weitzman (8) do not attribute this phenomenon to anything within the content of the test given but rather to something inherent in the position of the choices within the test items analyzed. This position factor is purported to hold true regardless of the type of test used. They contend that understanding the material presented in a

question is not the only factor at work in the selection of the correct answer. They believe that a subject does not always select his option on the basis of his fund of information alone but is influenced to some degree by these positional factors. The authors state that even though the position factor is found to be relatively small, it is statistically significant. Furthermore, they feel that its importance lies in its interaction with personality variables. Their hypothesis is essentially that both the first and the last items in a list are more outstanding than the middle two and therefore, since these inner choices are less noticeable they are less likely to be selected. Another possible explanation which they offer is that a person once having gone through the list without making a choice, rather than going through the list for a second time, is more likely to select the last choice. This however, is not sufficient enough to explain why the penultimate position is always found to be more difficult than the rest.

The most recent investigation of this general response pattern to multiple choice questions was undertaken by Clark (3) who collected data based on four standardized tests and found that these general positional preferences as they exist with five choice items are weak. A most important finding of this study reveals that when subjects are not pressed for time, these positional preferences are negligible, but when time is of the essence subjects were noted to

use the five positions in a slightly declining order from one to five. This was found to be significant at the two percent level.

Statement of the problem.---Previous studies dealing with the problem of position response set on objective type items were mainly concerned with either the effect of choice placement on item difficulty (8) or general positional preferences in terms of total number of responses (3,5,6,7). The present research attempts to combine both aspects by using more carefully controlled procedures than has been used in the foregoing studies. This study was conducted primarily to test the hypothesis that difficulty is effected by the position to which the correct response has been assigned. It was also a concern of this paper to investigate the possible tendency among college students to favor certain response positions to the neglect of others in four choice multiple choice questions.

Definition of terms:

A response set is here operationally defined as a deviation from chance at the five percent level of significance for the number of answers selected at each position. Consideration was given only to experimental differences which were significantly greater than the sampling differences that were likely to arise.

Difficulty is defined as the percentage of correct responses made out of the total number possible.

Hypotheses:

1. There is no difference in the difficulty level of an objective multiple choice item when the correct answer is moved from any one of the four possible positions to any other.

2. There is no difference in the frequency of selection of any position in a four option objective achievement test.

Method

Subjects.--For this experiment 434 students enrolled in various sections of the introductory course in psychology, given at the University of Massachusetts, served as subjects. The subjects from each class section were divided into four matched groups, equated on the basis of their test scores on the psychology mid-semester examination. This was achieved by sorting the test papers into four piles in an attempt to match individual scores as evenly as possible.

Stimulus materials.--The items selected for this study were taken from the final examination, given the preceding year, in the introductory course in psychology. Approximately two hundred test answer sheets from this examination were analyzed and a difficulty index, defined as the percentage of subjects passing an item, was computed for each of the one hundred items on the test. Once these indices were arrived at, the items were rearranged in the order of the magnitude of their difficulty; the easiest item being located in the first position and the most difficult item in the one hundredth. This was done to create a uniform set throughout the test and to increase the probability of eliciting a position preference as the items become more difficult. The test items in this examination were all four choice items and consisted of an incomplete statement which could be completed by having the examinee select one of the

four phrases following it. The following are samples of the questions which were used in the test.

1. Emotional behavior that is innate is made up of
 - (1) vocal responses.
 - (2) gestures.
 - (3) facial expressions.
 - (4) startle patterns.
2. Complex social needs are best related to
 - (1) primary incentives.
 - (2) secondary goals.
 - (3) complex human instincts.
 - (4) instrumental responses.

Procedure.--Since the literature cites instances where positional response sets have evidently occurred, it was recognized that not only may the student show a decided favoritism for certain positions but also that the test constructor may be affected by them in his placement of choices. If these positional factors do manifest themselves, then the probability of the student's pattern of preferences coinciding with that of the test constructor would necessarily be effected. To control for this possible bias, not only were an equal number of correct choices randomly scattered among the four positions but four different forms of the test were drawn up so that the correct choice for each item would appear in a different position on each of the forms.

The position that each alternative was assigned for each item was determined by following the scheme contrived by Mosier and Price (9) for arranging and randomizing correct choices and distractors. Twenty-four permutations of the numbers one through four were listed on separate slips

of paper which were then drawn singularly and at random. Each permutation was then assigned, to its final position in the test, in accordance with the order of appearance of its sequence number (see appendix). The correct answer was always written first with the plausible distractors following in sequence. Each successive item was then assigned the choice patterns in the order in which they occurred. Each choice pattern was used once before any pattern was repeated. For each successive form of the examination the entire cycle of alternatives was shifted over one position to the right, each form being derived from the previous one. Since the test was to be administered to a group of examinees, who differ in achievement in the basic psychology course, the equating of groups made it possible for the four forms of the test to be distributed equally among all levels of achievement. The assumption is that this method of distribution of forms among examinees virtually assures samples of both subjects and forms that are equivalent for all practical purposes.

The subjects were instructed to read each item very carefully and then to mark on their answer sheet that alternative which they have decided is most correct. They were told not to omit any items, and to guess when in doubt. Subjects were also informed that they would have sufficient time in which to complete all the items so that it would be unnecessary for them to rush. Such a procedure provides a

uniform response set, thereby minimizing individual differences in responding resulting from the operation of other sets.

All conditions for the collecting of data for the four forms were identical. Papers of those subjects who omitted any items or who responded more than once to an item were discarded.

Results

Distribution of test scores including mean scores for individual forms and for all forms combined are presented in Table 1.

Analysis of variance was used to determine whether there were any differences between the means of the four test groups who were each assigned a different form of the examination. The analysis reveals that the differences between means were not significant ($p > .20$) and that for all practical purposes the forms were not altered in any way as to affect the scores of the subjects from form to form.

When the total number of responses to each position (Table 2) were analyzed by means of the chi square test, for all forms combined, differences were found to be significant ($p < .05$). When the individual forms were analyzed, in terms of the total number of responses to each position all were found to be significant at the .05 level also, but on each form a different position was found to contribute most to the significance (see Fig. 1).

Application of the chi square technique to the total number of correct answers in each position (Table 3, Fig. 2) failed to yield significance for all forms combined or for any of the individual forms with the exception of Form C, which was significant at the .05 level.

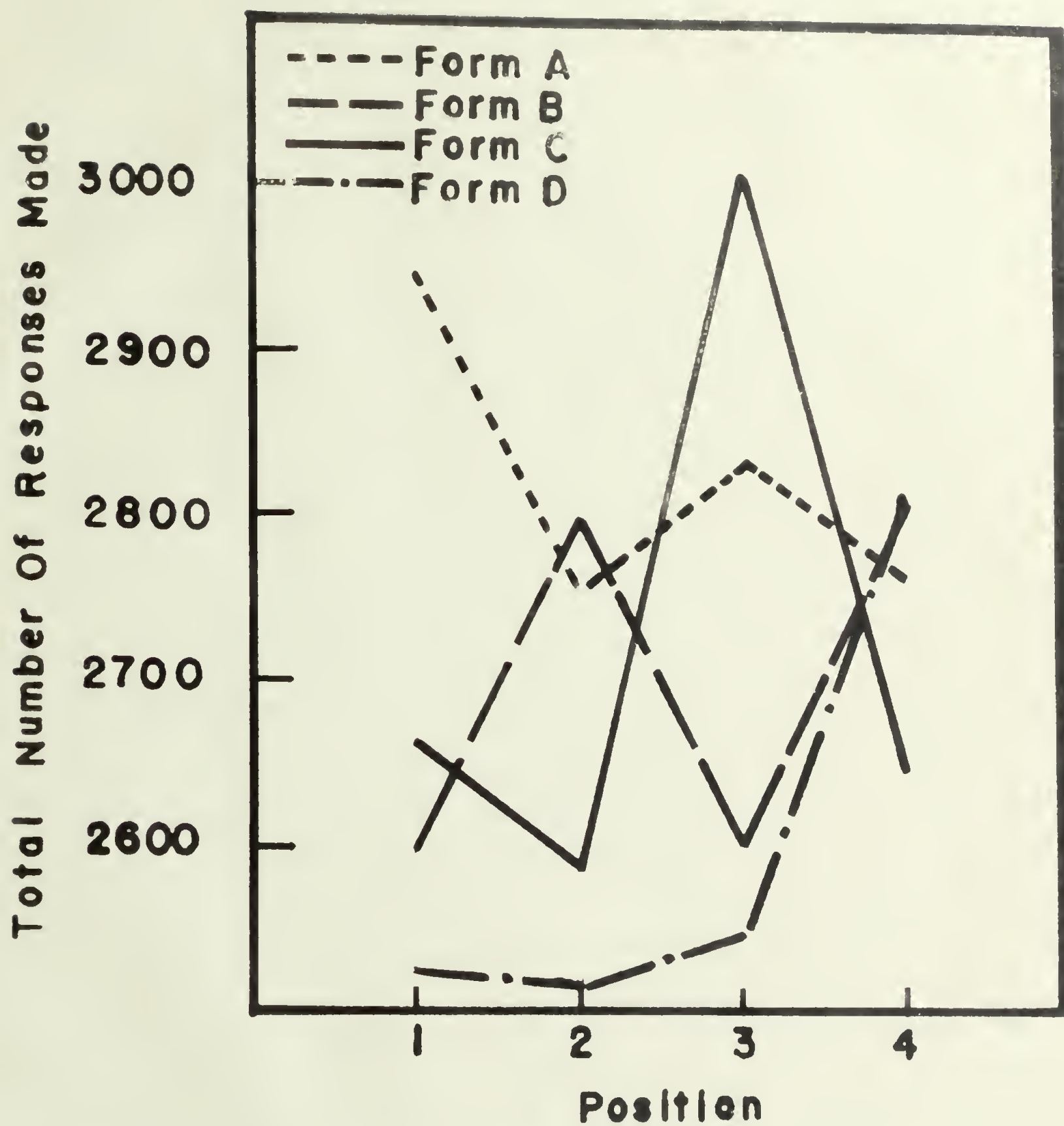
In order to determine whether the difficulty of any one

Table 1
Distribution of Scores
For All Forms

	Form A	Form E	Form C	Form D	Total
Total Score	7134	6811	7102	6606	27,653
Number in Group	113	108	109	104	434
Mean Score	63.13	63.06	65.16	63.52	63.72

Table 2
Total Number of Responses
To Each Position

	N	Position				Total
		1	2	3	4	
Form A	113	2946	2755	2833	2766	11,300
Form B	108	2597	2798	2601	2804	10,800
Form C	109	2662	2587	3005	2646	10,900
Form D	104	2525	2519	2547	2809	10,400
Total	434	10,730	10,659	10,986	11,025	43,400



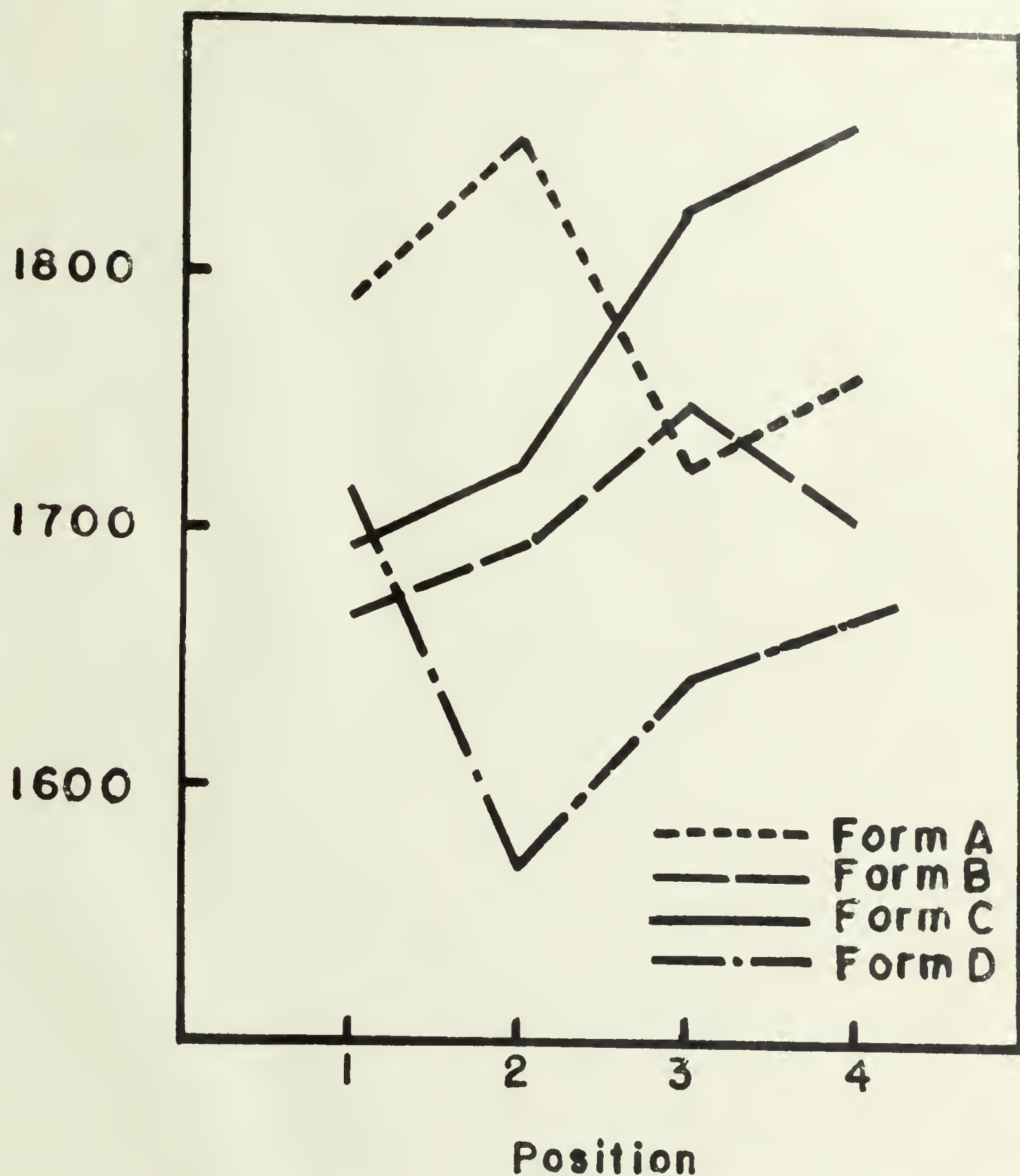
TOTAL NUMBER OF RESPONSES TO EACH POSITION

Fig. 1. Distribution of number of responses to each position for all forms.

Table 3
Total Number of Correct Answers
In Each Position

	N	Position				Total
		1	2	3	4	
Form A	113	1791	1856	1727	1760	7134
Form B	108	1664	1694	1749	1704	6811
Form C	109	1693	1723	1826	1860	7102
Form D	104	1717	1568	1640	1631	6606
Total	434	6865	6841	6942	7005	27,653

Total Number Of Correct Answers Made



TOTAL NUMBER OF CORRECT ANSWERS IN EACH POSITION

Fig. 2. Distribution of number of correct responses to each position for all forms.

position was significantly greater than any other, percentages of correct responses appearing in each position were computed (Table 4) for all forms combined and for individual forms alike. This index was computed by dividing the total number of observed responses made correctly in each position by the total number of possible correct answers in each position. Table 5 illustrates the percentage differences between the difficulty levels for the various positions and their critical ratios. The percentage difference between positions two and four was the only comparison found to be significant. Similar percentages were computed for all of those items in the test above the fifty percent level of difficulty (Table 6). These items, which numbered twenty in all, were analyzed in an effort to heighten the effect of a position response set, since it has been shown (4) that response sets are most apparent when items become more ambiguous or when they increase in level of difficulty. The percentage differences between the difficulty levels of the several positions for these items were not found to be statistically significant (Table 7).

For these same twenty items chi squares were also computed for the total number of responses to each position and for total number of correct responses in each position. Significance was not obtained on either of these tests.

Table 4
 Percentages of Correct Responses
 Appearing In Each Position
 On Each Form

	Position			
	1	2	3	4
Form A	63.40	65.70	61.13	62.30
Form B	61.63	62.74	64.78	63.11
Form C	62.13	63.23	67.01	68.06
Form D	66.04	60.31	63.08	64.65
Total	63.27	63.05	63.98	64.56

Table 5

Percentage Differences Between The Difficulty
Levels For The Various Positions And Their
Corresponding Critical Ratios

Correct Response Position	1	2	3
1			
2	.22 (.29)		
3	.71 (.96)	.93 (1.25)	
4	1.29 (1.69)	1.51 (2.04)	.58 (.79)

Critical Ratios are in parentheses.

Table 6

Percentages of Correct Responses Appearing
In Each Position For All Forms
(Items Over 50% Difficulty)

Position			
1	2	3	4
34.93	36.91	34.65	36.18

Table 7

Percentage Differences Between The Difficulty
Levels For The Various Positions And Their
Corresponding Critical Ratios
(Items Over 50% Difficulty)

Correct Response Position	1	2	3
1			
2	1.98 (1.35)		
3	.28 (.19)	2.26 (1.55)	
4	1.25 (.86)	.73 (.50)	1.53 (1.06)

Critical Ratios are in parentheses.

Discussion

The results of this research do not support the hypothesis, put forth by McNamara and Weitzman (3), that the difficulty level of a multiple choice test item is influenced by the location of the correct response. Furthermore, it is not even reasonable to assume that a tendency exists to select certain positions in a list of choices in preference to others. The findings do, however, indicate that the differences between observed and expected frequencies for total number of responses to each position over all forms were significant. Although the fourth position contributed most to the significance, it probably is of no great practical import, in that, on each individual form the position favored was a different one. The fact that the so-called position response set, as studied in the present investigation, does not carry over consistently from one form of the test to another suggests that there is no justification in referring to such behavior as response sets. In terms of the type of test materials used and the conditions under which they were administered, position preference responses seem to be of little consequence. Therefore, when multiple choice tests similar to this achievement test are used, these findings would indicate that position response sets should have little effect on test reliability and validity. Certainly there is no consistent bias from form to form as

would be implied by the McNamara and Weitzman (8) data.

These findings are in accord with Cronbach (5).

To understand the lack of agreement in results among total responses on each form it is necessary to consider the effects of the location of very plausible distractors. Even when the position of all alternatives is determined randomly, as was done in this experiment, it is quite possible that the most plausible distractors within any one grouping of test items will not be evenly distributed among the four possible positions. Since the sequence of correct and incorrect answers was systematically varied from form to form this may logically account for the differences in the distribution of total responses within forms. It is suggested that something other than a position response set is operating. Possibly the unequal attractiveness of distractors is what is contributing to the difference in frequency of response to each position, as indicated by the systematic shifting of preferences to a different position from form to form (see Table 2, Fig. 1). Another possible explanation might lie in the sequence effect of the position of the correct answer from item to item. Although there were an equal number of correct answers in each position throughout the test, the fact that an answer is in a certain position on one item may influence the test-taker to respond in a certain way on successive items. It should be understood that the arrangement of correct answers in a completely randomized

fashion, so that there is no repeating pattern of right answers, does not prevent an individual's response from being influenced by his previous responses. Actually the variance in results may be due to precisely such a set in the individual produced by the sequence effect.

Even if response sets were to operate, on the basis of probability theory alone, it would be negated by an equal distribution of correct choices over the items. Apparently, the test-taker could improve his score markedly, only if both he and the test constructor have the same marked preference for a particular position.

A knowledge of whether response sets exist or not is useful inasmuch as any influence which tends to cause subjects to mark one response more often than another is a factor which influences the validity of the right minus $1/4$ wrongs, method of scoring. This method is based upon the assumption that the probability is 1 out of 4 that a correct guess will be made on a given item.

Although the present study has been mainly concerned with the achievement test, it is quite possible that other types of tests employing multiple choice items may actually be subject to a position response set. Personality tests and other inventories which employ fixed categories of responses, have been known to reveal individual tendencies or sets which may be due in some respect to this position factor (1,2,6,7,10).

The crucial question for an understanding of the problem of response set is the extent to which the set is stable or fixed. It is apparent from the data obtained in this study that the position effect is not constant since it shifts from form to form. Since response sets have no opportunity to show themselves when the subject gets most items correct they should be apparent, if they exist at all, on the more difficult items. Analysis of the data for items over the fifty percent level of difficulty revealed no significant differences whatsoever. In light of these findings and those of Cronbach (5), it may be assumed that multiple choice items on objective type achievement tests are generally free from position preference response sets.

Summary

The present investigation tested the hypothesis that the difficulty level of an objective type multiple choice item is a function of the position to which the correct response has been assigned. In this study one hundred multiple choice achievement test items, given as part of the final examination in the introductory psychology course, were used. Subjects were equated into four groups and four different test forms were devised, with the correct choice for each item appearing in a different position on each of the forms. Four hundred and thirty-four undergraduate students at the University of Massachusetts served as subjects.

The results indicate that a position preference hypothesis is untenable. It is suggested that the position of the most plausible distractor may more logically account for the significant position effect than does any position preference. The research lends no support to the McNamara and Weitzman (8) hypothesis that the next to last position is most difficult. Over all forms, the penultimate position was actually found to be among the least difficult.

In view of the non-supportive results obtained in this study and the low reliability of position preferences reported by Cronbach (5), it appears that position preference is not a significant source of invalidity in multiple choice achievement tests. A more promising line of research might

be to investigate the sequential effects from item to item or to attempt to get at a method for equalizing the attractiveness of distractors.

References

1. Berg, I. A. The reliability of extreme position response sets in two tests. J. Psychol., 1953, 36, 3-9.
2. Berg, I. A. and Collier, J. S. Personality and group differences in extreme response sets. Educ. psychol. Measmt., 1953, 13, 164-169.
3. Clark, E. L. General response pattern to five choice items. J. educ. Psychol., 1956, 47, 110-117.
4. Cronbach, L. J. Response sets and test validity. Educ. psychol. Measmt., 1946, 6, 475-493.
5. Cronbach, L. J. Further evidence on response sets and test design. Educ. psychol. Measmt., 1950, 10, 3-31.
6. Mathews, C. O. The effect of position of printed response words upon children's answers to questions in two-response types of tests. J. educ. Psychol., 1927, 18, 445-457.
7. Mathews, C. O. The effect of the order of printed response words on an interest questionnaire. J. educ. Psychol., 1929, 20, 128-134.
8. McNamara, W. J. and Seitzman, E. The effect of choice placement on the difficulty of multiple-choice questions. J. educ. Psychol., 1946, 36, 103-113.
9. Mosier, C. I. and Price, H. G. The arrangement of choices in m-c questions and a scheme for randomizing choices. Educ. psychol. Measmt., 1945, 5, 379-382.
10. Rapaport, G. M. and Berg, I. A. Response sets in a multiple choice test. Educ. psychol. Measmt., 1955, 15, 58-62.

Appendix

Instructions Administered to Groups

Name _____

Section _____

Psychology 26

FINAL EXAMINATION

DO NOT OPEN THIS EXAMINATION BOOKLET UNTIL
YOU ARE TOLD TO DO SO

1. Write your name in the appropriate blank of your ANSWER SHEET.
2. Write your instructor's name and your section number directly under your name.
3. Write your name, your instructor's name, and your section number on all essay answer sheets.
4. Read each question CAREFULLY and when you have decided which alternative is MOST correct, with your special pencil blacken the space that corresponds to your choice. Do not make more than one mark per question; the question will be marked wrong automatically if there are any other marks in the other blanks.
5. Do not make any unnecessary marks on the answer sheet. If you make a mistake, completely erase the black mark.
6. Be sure that the number of the question you are answering always corresponds to the number on the answer sheet.
7. Answer every question. If you are not sure of an answer, make the best choice you can. When in doubt GUESS.
8. Try not to spend more than one hour (approximately) on this part. This time should be sufficient to complete all of the multiple choice items.
9. When you have finished your examination, check your work, then put all the materials back into this booklet and give them to your instructor.
10. Once you have left the examination room, do not come back for books or wraps until the examination is over.

Randomization For Choice Patterns*

1. 4213	26. 2143	51. 1423	76. 2134
2. 4312	27. 3241	52. 1243	77. 3214
3. 2431	28. 3214	53. 2341	78. 1243
4. 1432	29. 2341	54. 2413	79. 1342
5. 3412	30. 3142	55. 1234	80. 2431
6. 4321	31. 4312	56. 2134	81. 1234
7. 4231	32. 1432	57. 2314	82. 4213
8. 1342	33. 1243	58. 2143	83. 1324
9. 3421	34. 4321	59. 3142	84. 3142
10. 3214	35. 3421	60. 1342	85. 2143
11. 1243	36. 1234	61. 3412	86. 4123
12. 1423	37. 2413	62. 4123	87. 1423
13. 2314	38. 1342	63. 2431	88. 4132
14. 2134	39. 4132	64. 4132	89. 4231
15. 4123	40. 2314	65. 2143	90. 2341
16. 2143	41. 1423	66. 1324	91. 3124
17. 3241	42. 2134	67. 4213	92. 3421
18. 1324	43. 4231	68. 3241	93. 4321
19. 4132	44. 2431	69. 4321	94. 3241
20. 3124	45. 3412	70. 1432	95. 4312
21. 2341	46. 4213	71. 4312	96. 3412
22. 2413	47. 3124	72. 3421	97. 1432
23. 3142	48. 4123	73. 3124	98. 4321
24. 1234	49. 3412	74. 3214	99. 2314
25. 1234	50. 1324	75. 4231	100. 2413

*The number "1" in all the above patterns indicates the position of the correct answer.

Acknowledgments

The writer wishes to express his sincere appreciation to Dr. Jerome L. Myers, Chairman of his thesis committee, who most generously gave of his time and was throughout the completion of this investigation a constant source of guidance and encouragement.

Gratitude is also extended to Miss Barbara Goldberg for her assistance in the tabulation of the test material data and to all the members of the psychology department teaching staff, without whose cooperation this study could not have been carried out.

Approved by:

J. F. Myers

William F. Field

Alb. Trumbull

