Florida Journal of Educational Research, 1971, 13

COMPARISON OF THREE MATCHING ITEM FORMATS

John Follman, Raymond Urbanke and **Wade Burley**

University of South Florida

SUMMARY

Three matching item test formats were compared empirically. The random order test format produced performance significantly superior to both the homogeneity of meaning test format and the one overall group test format. It was concluded: that the format of matching tests influences performance; that random ordering of matching items produces highest performance; that instructors should determine that both they and their students have denotative knowledge of the meaning of the so-called key words used in essay questions; and that matching assignments should not involve too large a number of stimuli and responses.

Authors of measurement books strongly recommend that matching test items be grouped homogeneously (Adams, 1964; Ahmann & Glock, 1967; Green, 1963; Payne, 1968; Thorndike & Hagen, 1955). Homogeneous grouping represents similarity of stimuli, similarity of responses, and similarity across stimuli and responses. Similarity means the degree of relatedness within particular domains such as dates, names, (Payne, 1968). None of these authors, however, cites empirical evidence of the effects of homogeneous item grouping or other formats of item grouping on test performance.

The purpose of this study was to compare three test formats of presentation of matching items. The formats were "homogeneity of meaning," "random order," and "one overall group." It was hypothesized that there would be no significant differences in performances between: the homogeneity of meaning and random order formats; the homogeneity of meaning and one group of formats; and the random order and one group formats.

Experimental Materials

Weidemann (1941) analyzed essay examination questions and prepared a list of 11 kinds of essay questions from simple to complex; what, who, when, which, where, list, outline, describe, contrast, compare, explain, discuss, develop, summarize, evaluate.

Bird and Bird (1945) compiled a list of words, appearing frequently in essay questions, which purportedly are critical and/or crucial to students' understanding of the denotative meaning of the essay questions. Nearly all of the Weidemann (1964) kinds of essay questions and/or words were included in this list. Morgan and Deese (1957) revised somewhat the definitions of these key words. Twenty of the 21 key words were used in the current study as the experimental medium. The word "review" was not included. The definitions presented in Morgan and Deese (1957) were edited so that the differences between definitions were sharpened and the number of key or distractor words included as synonyms in the definition of other words were reduced to a minimum.

Subjects

The subjects were 60 undergraduates, predominately sophomores and juniors enrolled in an educational psychology course at the University of South Florida during the third quarter of the 1967-68 academic year.

The students were randomly assigned to the three matching item test format groups. The task was to match each key word with its definition. There was no time limit.

In the **homogeneity of meaning** test format there were six groups of matching items grouped by the investigators to be relatively homogeneous with respect to denotative meaning of the key words. The six groups of key words ranged in number from two to five. In addition each group included one distractor word so that the key word groups actually ranged from three to six words for a grand total of 20 key words and six distractor words.

In the **random order** test format the 20 key words and five distractor words were presented in five groups of matching items in random order of presentation as well as random order within groups both for the key words and their definitions. Again one distractor word was included in each key word group.

In the **one overall group** test format the 20 key words and five distractor words were presented in one overall group as were their definitions. The definitions were randomly ordered as were the key and distractor words.

The stimulus materials each representing one of the three treatments are given in Table 1.

Table 1

Homogeneity of Meaning

Establish that something is true by citing factual evidence or giving clear logical reasons.	a. prove
Give reasons for decisions or conclusions, taking pains to be convincing.	b. justify c. confirm
Give a drawing, chart, plan, or graphic answer.	d. interpret
Translate, give examples of, solve, or comment on a subject usually giving your judgment about it.	e. diagram f. chart
Use a figure, picture, or concrete example to clarify a problem.	g. illustration
Look for qualities or attributes that resemble each other. Emphasize similarities but in some cases also mention differences.	h. criticize i. differentiate
Carefully appraise the problem, citing both advantages and limitations. Emphasize the view of authorities and to a lesser degree your view.	j. evaluate k. contrast
Express your judgment about the merit or truth of the factors or views mentioned. Discuss limitations and good points.	l. compare
Stress the dissimilarities, differences, or unlikenesses of things, qualities, events, or problems.	
Clarify, and spell out the material you present.	m. discuss
Examine, analyze carefully, and give reasons pro and con. Be complete and give details.	n. explain o. define
Give clear, concise and authoritative meanings. Give the limits but not the details. Show how the thing differs from the things in other classes.	p. characterize
Organize under main points and subordinate points omitting minor details and stressing the arrangement or classification of things.	q. outline r. state
Present the main points in brief, clear sequence, usually omitting details, illustrations, or examples.	s. summarize t. list u. count
Write in numbered form, giving points concisely one by one.	v. enumerate
Give the main points or facts in condensed form omitting details and illustrations.	
Write an itemized series of concise statements.	
In sequential form describe progress, development, or historical events from some point or origin.	w. narrate
Show how things are connected with each other or how one causes another, correlates with another or is like another. Recount, characterize, sketch, associate in sequence or story form.	x. describe y. trace z. relate

Random Assignment

Give a drawing, chart, plan, or graphic answer.	a. confirm
Give the main points or facts in condensed form omitting details and illustrations.	b. contrast
Stress the dissimilarities, differences, or unlikenessess of things, qualities, events, or problems.	c. summarize
Carefully appraise the problem, citing both advantages and limitations. Emphasize the view of authorities and to a lesser degree your view.	d. diagram
	e. evaluate
Clarify, and spell out the material you present.	f. list
Write an itemized series of concise statements.	g. characterize
Translate, give examples of, solve, or comment on a subject usually giving your judgment about it.	h. interpret
Express your judgment about the merit or truth of the factors or views mentioned. Discuss limitations and good points.	i. explain
	j. criticize
In sequential form describe progress, development, or historical events from some point or origin.	k. outline
Organize under main points and subordinate points omitting minor details and stressing the arrangement or classification of things.	l. relate
Look for qualities or attributes that resemble each other. Emphasize similarities but in some cases also mention differences.	m. compare
Show how things are connected with each other or how one causes another, correlates with another or is like another.	n. count
	o. trace
Give clear, concise and authoritative meanings. Give the limits but not the details. Show how the thing differs from the things in other classes.	p. discuss
Present the mainpoints in brief, clear sequence, usually omitting details, illustrations, or examples.	q. prove
Examine, analyze carefully, and give reasons pro and con. Be complete and give details.	r. state
Establish that something is true by citing factual evidence or giving clear logical reasons.	s. narrate
	t. define
Give reasons for decisions or conclusions, taking pains to be convincing.	u. describe
Use a figure, picture, or concrete example to clarify a program.	v. enumerate
Write in numbered form giving points concisely one by one.	w. illustrate
Recount, characterize, sketch, or associate in sequence or story form	x. justify
	y. differentiate

One Overall Group

___1. Carefully appraise the problem, citing both advantages and limitations. Emphasize the view of authorities and to a less degree your view.

___2. Present the main points in brief, clear sequence, usually omitting details, illustrations, or examples

___3. Establish that something is true by citing factual evidence or giving clear logical reasons.

___4. Translate, give examples of, solve, or comment on a subject usually giving your judgment about it.

___5. Recount, characterize, sketch, or associate in sequence or story form.

___6. Show how things are connected with each other or how one causes another, correlates with another or is like another.

___7. Give clear, concise and authoritative meanings. Give the limits but not the details. Show how the thing differs from the things in other classes.

___8. Clarify, and spell out the material you present.

___9. Stress the dissimilarities, differences, or unlikenesses of things, qualities, events, or problems.

___10. Express your judgment about the merit or truth of the factors or views mentioned. Discuss limitations and good points.

___11. Give a drawing, chart, plan or graphic answer.

___12. Write in numbered form, giving points concisely one by one.

___13. Give the main points or facts in condensed form omitting details and illustrations.

___14. In sequential form describe progress, development, or historical events from some point or origin.

___15. Organize under main points and subordinate points omitting minor details and stressing the arrangement or classification of things.

___16. Give reasons for decisions or conclusions, taking pains to be convincing.

___17. Look for qualities or attributes that resemble each other. Emphasize similarities but in some cases also mention differences.

___18. Use a figure, picture, or concrete example to clarify a problem.

___19. Write an itemized series of concise statements.

___20. Examine, analyze carefully, and give reasons pro and con. Be complete and give details.

a. justify
b. interpret
c. prove
d. compare
e. relate
f. confirm
g. criticize
h. illustrate
i. outline
j. define
k. explain
l. count
m. characterize
n. evaluate
o. trace
p. describe
q. differentiate
r. enumerate
s. diagram
t. contrast
u. state
v. list
w. summarize
x. discuss
y. narrate

Statistical Analysis

Kuder-Richardson Formula (KR) 20 and 21, and Spearman-Brown correct split half odd even reliability estimates were computed on total test score for each of the three test formats. Means, medians, variances and "t" tests were calculated to determine if there were significant differences in performance between: the homogeneity of meaning and random order format test groups; the homogeneity of meaning test and one overall group test formats; and the random order test and the one overall group test formats.

The proportion of Ss in the lower 27% on total test score which correctly matched each item was subtracted from the proportion of Ss in the upper 27% on total test score which correctly matched that item for all items for each format. The mean discrimination index was then computed for each format.

Results

Table 2 indicates the N's, means, medians, variances, KR 20 and KR 21 and odd-even reliability estimates, and mean discrimination indices for the homogeneity of meaning, random order, and one group matching test formats.

The variances for the homogeneity of meaning group and the random order matching test formats were significantly different at the .01 level. Therefore the separate variance formula was used. The t-ratio was 7.65, significant at the .001 level. The random order format is apparently easier for students than the homogeneity of meaning format, probably because the homogeneity of similarity of meaning of the various stimuli and responses produces greater inhibition in stimuli, responses, and the matching of stimuli and responses.

The t-ratio computed from the pooled variance formula was 3.75, significant at the .001 level, between the homogeneity of meaning test and the one overall group matching test formats. The homogeneity of meaning format is apparently somewhat easier than the one overall group format probably because it is very difficult for a student to retain and manipulate without error 25 key and distractor words, and 20 definitions.

The variances for the random order group, and the one overall group matching test formats were significantly different at the .05 level. Therefore the separate variance formula was used. The t-ratio was 13.28, significant at the .001 level. The random order format is apparently much easier than the one overall group format both because there is relatively little inhibition due to similarity of stimuli and responses, and also because it is difficult to keep track of and manipulate without error the 25 key and distractor words and the 20 definitions.

The KR 21 reliability estimates of .511 for the homogeneity of meaning matching test format, and .609 for the randomly ordered matching test format indicate similar, moderate levels of strength. The estimates of .319 for the one overall group matching test format suggests low realibility. This low reliability probably accounted in part at least for the low mean performance on this format. This may well reflect the difficulty for the students of attempting to sort out and keep track of the 25 key and distractor words, and the 20 definitions, a challenging assignment.

Table 2

Ns, Means, Medians, Variances, KR 20 and KR 21, and Odd-Even Reliability Estimates and Mean Difficulty Indices for the Homogeneity of Meaning, Random Order and One Group Matching Test Formats

	N	X	Median	Variance	KR 20	KR 21	OE*	X DI
Homogeneity	20	12.95	13	8.89	.681	.511	.799	.39
Random	18	18.83	19	2.62	.689	.609	.717	.11**
One Group	22	9.68	9	7.18	.455	.319	.419	.32

*Spearman-Brown Corrected

**Upper 50% vs. lower 50% used rather than 27% - 27%.

Inspection of the item discrimination indices for the homogeneity of meaning format and the one group format indicated little overlap between the items discriminating between those who were in the upper 27% and those in the lower 27% on the one hand, and the words most frequently matched incorrectly on the

other hand. This may be interpreted to mean that successful matching is determined more by the format than by knowledge of the definitions of the words. For the random format so few items were incorrectly matched, the distribution was so negatively skewed, that the difficulty estimates and discrimination indices were meaningless.

The random format and the homogeneity of meaning format were inspected for evidence that the matching performance might have been influenced substantially by the number of matchings made per group. In the random format the total number of misses for each group of four across all Ss were low (1, 8, 2, 6, 2) with the frequency of misses for the word most incorrectly matched being only three. In the homogeneity of meaning format the number of matchings to be made per group and their corresponding number of incorrect matches were, 2-3, 3-21, 4-21, 3-23, 5-41, and 3-21. It can be seen that as the number of matchings to be made increases so do the errors as would be expected, particularly when performance on the one overall test format group is considered. The first group of two matchings had a total of only three errors and the fifth group of five matchings had 41 errors but the other four groups were generally consistent in proportion of errors made to number of matchings made. When the words missed are noted it is apparent that most are words also missed most frequently in the one overall group format, and also in the alphabetical format and to a limited extent in the random format in another study by Follman, Lowe, and Burley (1968).

The words most frequently matched incorrectly in the homogeneity of meaning format were: describe, state, define, list, discuss, enumerate, and diagram. The words most frequently incorrectly matched in the one group format were: describe, criticize, define, state, enumerate, discuss, and evaluate. Since the highest frequency of incorrect responses per word was only three for the random format the words most frequently missed are not considered meaningful and will not be reported. It is suggested that the words missed most frequently: describe, state, discuss, and enumerate, among others, be defined explicitly for their students by instructors anticipating using them in examination questions.

Finally, the superior performance associated with the random format over performance on either of the other two formats provides evidence that is consistent with the results of studies by

Marcus (1963) and Wevrick (1962) which may be interpreted to suggest that if the position of the correct alternative in multiple choice options is randomized, response set will be minimized.

Conclusions

One conclusion is that matching item test format influences test performance. It may be speculated that performance on some matching tests may be more a function of the format than knowledge of the content. The findings of this study may also be interpreted as justification for the considerable coverage given by most measurement text authors to the writings of "good" matching items.

Another conclusion is that the random order format produces higher test performance. It might be contended that the spread of responses should be less with this format than with the homogeneity of meaning, one group, or perhaps other formats. However a more realistic argument would appear to be that in actual testing situations the spread that is desired should be a function of knowledge of the appropriate content rather than the format of the presentation of the items.

The third conclusion is that the assignment of matching 20 definitions with 25 key and distractor words is too difficult and groups with smaller numbers of stimuli and responses should be used.

The final conclusion is that, insofar as the so called key words, particularly describe, state, discuss, enumerate, are actually used in essay questions, teachers should insure that both they and their students have denotative knowledge of the words' meaning.

References

Adams, G. S. *Measurement and evaluation.* New York: Holt, Rinehart & Winston, 1964.

Ahmann, J. S., and Glock, M. *Evaluating pupil growth.* Third Edition. Boston: Allyn & Bacon, Inc., 1967.

Bird, C., & Bird, D. M. *Learning more by effective study.* New York: Appleton-Century-Crofts, 1945.

Follman, J. C., Lowe, A. J., and Burley, W. W. Relationship of two matching item test formats to the Nelson-Denny Reading Test. Unpublished study, 1968.

Green, J. A. *Teacher made tests.* New York: Harper and Row. 1963.

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

Morgan, C. T., & Deese, J. *How to study.* New York: McGraw-Hill Book Company, Inc., 1957.

Payne, D. A. *The specification and measurement of learning outcomes.* Waltham, Massachusetts: Blaisdell Publishing Company, 1968.

Thorndike, R. L. & Hagen, E. *Measurement and evaluation in psychology and education.* New York: John Wiley & Sons. 1955.

Weidmann, C. C. Review of essay test studies. *Journal of Higher Education*, 12, 1941, 41-44.

Wevrick, L. Response set on a multiple-choice test. *Educational and Psychological Measurement*, 22, 1962, 533-538.