.OCUMENT RESUME

ED 078 001 TM 002 813

AUTHOR	Price, Jay R.; And Others
TITLE	Construct Validity of Test Items Measuring Acquisition of Information from Line Graphs.
PUB DATE	73
NOTE	15p.; Paper presented at American Educational Research Association Meeting (New Orleans, Louisiana, February 25-March 1, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Classification; *Graphs; *Item Analysis; Research Methodology; Technical Reports; Test Construction; *Test Validity

ABSTRACT

Research on the effectiveness of graphical displays for information acquisition and retention lacks a system for classifying graph information and generating test items to assess learning. The purpose of this study was to validate a system based on two types of information and three types of informational units. Results of an analysis of variance indicated differences in learning predictable from the classification system; however, a multitrait-multimethod matrix analysis failed to provide evidence of trait validity for the system's informational constructs. In light of these results, a graph information processing strategy was proposed in which subjects utilize data point information. (Author)

33-08

Title

Construct Validity of Test Items Measuring Acquisition of Information from Line Graphs

Authors and Institutional Affiliation

Jay R. Price[1], University of Delaware

Victor R. Martuza, University of Delaware

James H. Crouse, University of Delaware

Footnotes Pertaining to Identity of Authors

1. Requests for reprints should be sent to Jay R. Price, College of Education, University of Delaware, Newark, Delaware 19711.

Mail Proof To:

Jay R. Price

College of Education

University of Delaware

Newark, Delaware 19711

AERA New Orleans '73

Construct Validity of Test Items Measuring Acquisition of Information from Line Graphs

Abstract

Research on the effectiveness of graphical displays for information acquisition and retention lacks a system for classifying graph information and generating test items to assess learning. The purpose of this study was to validate a system based on two types of information and three types of informational units. Results of an analysis of variance indicated differences in learning predictable from the classification system; however, a multitrait-multimethod matrix analysis failed to provide evidence of trait validity for the system's informational constructs. In light of these results, a graph information processing strategy was proposed in which subjects utilize data point information.

Construct Validity of Test Items Measuring Acquisition of Information from Line Graphs

The present study deals with the acquisition and retention of quantitative information from a line graph stimulus. While the acquisition of quantitative information from graphical displays is an important component of school learning, the processes involved in such situations have been studied only infrequently, (cf., Washburne, 1927; Schutz, 1961). The present study is particularly concerned with three aspects of learning from a line graph stimulus; (a) the nature of the informational unit(s) processed by subjects instructed to learn the information in the graph, (b) the relationship between the number of informational units upon which a test item is based and accuracy of subject performance on that item, and (c) the relationship between study time and acquisition of information from the graph.

In attempting to measure the acquisition of information from a line graph stimulus, the first question which arises concerns the nature of the informational units processed by the subject. A logical distinction exists between point and slope information. In a line graph, a unit of point information is the value of the dependent variable associated with a specific level of the independent variable; a unit of slope information is the change in value of the dependent variable per unit change in the independent variable associated with a specific, contiguous set of independent variable levels. The question of immediate interest is whether this logical distinction is a meaningful psychological distinction; i.e., when instructed to learn the information in a line graph, do subjects encode point and/or slope information? If subjects do, in fact, store point and slope information independently, then point and slope information can be viewed as informational constructs in much the same way that personality constructs are viewed; thus, it should be possible to validate items measuring these informational constructs by means of multitrait-multimethod methodology (Campbell and Fiske, 1959).

The second question of interest concerns the relationship between the number of informational units required for correct performance on items at recall and accuracy of subject performance on these items. Studies by Schutz (1961) and Washburne (1927) are tangentially related to this question, but because of differences in procedure, task instructions, and type of item presentation format, the studies do not lead directly to expectations for the present experiment. However, it would seem that the greater the number of informational units required by an item at recall, regardless of the type of unit involved, the poorer performance should be on the item.

The third question of interest concerns the effects of study time on information acquisition. The purpose here was to extend the research on study time into the area of learning quantitative information from graphical materials. It was expected, as most studies have shown, that increased study time would result in greater acquisition. Of greatest interest, however, were the possible interactions of study time with the type of informational units and with the number of informational units which were required for successful performance on the test items at recall.

Method

Subjects. Thirty-six undergraduate education student volunteers served as subjects in this experiment.

Materials. A multiple line graph was constructed in which the average value per share of stock for each of three fictitious companies was plotted for each of five successive years. Each of the three lines (one per company) was generated randomly, subject to the following constraints: (a) one line would show an increasing trend, (b) the second line would show a decreasing trend, and (c) the third would show random fluctuations. To generate the data points for the first two of these lines, the data point values were randomly sampled from the following five strings of digits: 0-5, 1-6, 2-7, 3-8, and 4-9. For the increasing trend line, the first digit was randomly selected from the 0-5 interval. The next four digits were randomly selected from the four succeeding digit strings. For the decreasing trend line, the first digit was randomly selected from the 4-9 interval. The next four were randomly selected from the remaining intervals in sequence. The five values for the third line were randomly selected from the 0-9 range subject to the restriction that there would be exactly one intersection or crossover of lines in the left, center, and right thirds of the graph.

The criterion test consisted of six subtests of eight propositions each. Three subtests were based on point information; the rest on slope information. Within each information type, the three subjects were based on a single unit of information, two units arranged vertically (i.e., the price of stock for two companies during the same year), and two units within the same line (i.e., the price of a single company's stock for two separate years) respectively. Following the lead of Anderson (1972), Bormuth (1970), and Cronbach (1971), basic sentence frames were formed for each item type (See Table I) and rules were established to generate the items in each cell.

Table I about here

For example, the rules for the point items based on a single unit of information are listed below:

1. Company names for the eight items were selected randomly with the restriction that each company name was used at least twice and no more than three times.
2. The year values for the eight items were chosen randomly with the restriction that each year value was used at least once and no more than twice.
3. The comparative (greater than-less than) was assigned randomly to the items so that each appeared in four items of the subtest.
4. Within the four items containing the 'greater than' comparative, the truth value was randomly assigned such that two propositions would be true and two would be false. The same procedure was used for the four 'less than' comparative items.
5. For each item, the set of stock values which would satisfy the truth value for that item was determined and one element of the set was randomly selected for inclusion in the item.

It is apparent from the above rules that items within each subtest were balanced for wording of comparatives (e.g., greater than-less than, more rapidly-less rapidly, increased-decreased) and truth value. With respect to wording of comparatives, a number of researchers (e.g., Clark, 1970, Trabasso, 1970) have shown that positive and negative wording of test items impose different information processing requirements on subjects with resulting differences in performance levels. These results as well as those on acquiescent responding suggested that items should be balanced for comparative wording and truth value so that comparisons of interest would not be differentially contaminated by differences in responding.

Analogous procedures were used for generating each of the five remaining item types. The items were then randomly ordered over the test as a whole, subject to constraints necessary for guaranteeing that the distribution of the various item characteristics described above would be even across the test as a whole.

The graph and test items were reproduced on standard 8 1/2" x 11" sheets of paper and bound in a seven page test booklet. A cover sheet for subject identification was followed by the graph. A blank sheet followed the graph and separated it from the three pages of test propositions to prevent the subjects from seeing the graph at test time. A final cover sheet completed the test booklet.

Procedure. The subjects were randomly assigned in equal numbers to the two and eight minute study time conditions. Following distribution of the materials, instructions were read to the subjects which (a) indicated the purpose of the study, (b) specified both the study time and test time limits, (c) informed them that the graph could not be used as a reference once the prescribed study time had elapsed and (d) instructed them to answer all items. Subjects were told they had up to 40 minutes to complete the test items. As it turned out, no one required more than 25 minutes to complete the test items.

Results

The number of correct responses per item type was determined for each subject. These data were then analyzed as a one-between, three-within factorial analysis of variance. The between factor was study time and the within factors were information type, number of informational units, and wording of logical opposite pairs. Table II contains the means and standard deviations for this analysis.

Table II about here

All four main effects were significant while none of the interactions was significant. The mean score in the eight minute study condition was higher than the mean in the two minute condition, $F = 10.90$; $df = 1/34$, $p < .01$. The mean score on point information items was significantly higher than the mean on slope information items $F = 6.18$, conservative $df = 1/34$, $p < .02$. Scheffe tests on the three information unit means indicated that the mean of single unit items was higher than the weighted means of the two unit within occasion and two unit within group items ($p < .01$); however, the means of the latter two item types were not significantly different from each other ($p > .05$). The mean performance on items stated positively (greater than, increase, more rapidly) was significantly higher than mean performance on items stated negatively, $F = 6.16$, conservative $df = 1/34$, $p < .02$.

To assess the relationship between performance nformation types and number of data points required to answer an item successfully, the six subtest means (information type X number of units) were analyzed as a one-between, one-within analysis of variance (time X subtest). The two main effects were significant, the interaction was not. The number of data points and subtest means as well as the significant comparisons by the Newman-Keuls procedure are contained in Table III. This analysis indicated that only the mean of the point-single unit test differed significantly from the means of the slope-within occasion and slope-within group tests.

Table III about here

In order to assess possible effects of response sets, the data were reanalyzed with study time, logical opposite pairs, and truth value as the independent variables. The only significant results were those main effects associated with study time and logical opposite wording. The fact that all interaction effects were nonsignificant seems to rule out acquiescence as a possible explanation for the results obtained in the initial analysis discussed above.

Table IV contains the multitrait-multimethod matrix with number of informational units representing the methods, and point and slope information being the possible constructs. Correlation coefficients appearing in the table have been corrected for attenuation. The overall pattern of coefficients in the matrix does not support our hypothesis that the point and slope items included in this criterion test measure two distinct informational constructs.

Table IV about here

Discussion

The results of the initial analysis indicated significant main effects for study time, wording, number of informational units, and informational types. The effect of informational types suggested that the point-slope dichotomy was a meaningful distinction; however, the multitrait-multimethod matrix failed to support this distinction: performance on the various point and slope subtests predicted performance on subtests both within and between these two informational constructs.

An explanation for the disparate results of these two analyses may lie in the kind of information subjects encoded and/or retrieved under the experimental instructions and conditions of this study. It is possible that subjects did not use slope information as defined in this study but instead used only data point information. To answer slope items, subjects recalled point information and then constructed slope information from the recalled points. The reasoning which follows supports this conclusion.

Slope items are apparently more difficult than point items. If slope performance is a functionof a subject's recall of data points, then an increase in the number of data points needed for successful performance should be accompanied by a decrease in performance level. From Table III, it is apparent that this inverse relationship exists; subjects' scores tend to decrease as the number of data points increases.

Consequently, it appears that the amount of data point information may be a more important factor than informational type in determining a subject's performance level given the proposed information processing strategy. However, the present findings do not rule out the possibility that under other experimental instructions and conditions, subjects would encode slope information. If this were the case, then the present multitrait-multimethod methodology seems suitable for providing evidence of the encoding of slope information and the validity of the slope informational construct.

References

Anderson, R.C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Bormuth, J. On the theory of achievement test items. Chicago: The University of Chicago Press, 1970.

Campbell, D.T., & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Clark, H.H. How we understand negation. Paper presented at the Workshop on Cognitive Organization and Psychological Processes, Huntington Beach, California, August, 1970.

Cronbach, L.J. Test validation. In R.L. Thorndike (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1971.

Schultz, H.G. An evaluation of formats for graphic trend displays--Experiment II. Human Factors, 1961, 99-107.

Trabasso, T., Rollins, H., & Shaughnessy, E. Storage and verification stages in processing concepts. Cognitive Psychology, 1971, 2, 239-289.

Washburne, J.N. An experimental study of various graphic, tabular, and textural methods of presenting quantitative material. Journal of Educational Psychology, 1927, 18, 361-376, 465-476.

Table I

Test Item Frames--Information Type and Number of Informational Units

	Single Unit	Two Units--Within Occasion	Two Units--Within Group
Point	The average value per share of (name) stock was (±comparative) than (value) during the year (year).	The average value per share of (name) stock was (±comparative) than the average value of (name) stock during the year (year).	The average value per share of (name) stock was (±comparative) in (year) than in (year).
Slope	The average value per share of (name) stock (±verb) during the period (years).	The average value per share of (name) stock changed (±comparative) rapidly than the value of (name) stock during the period (years).	The average value per share of (name) stock changed (±comparative) rapidly during the period (years) than during the period (years).

Table II

Means and Standard Deviations--Study Time X Information Type X Units of Information X Wording

		Point						Slope					
		1 unit		2 w. group		2 w. occasion		1 unit		2 w. group		2 w. occasion	
		+	-	+	-	+	-	+	-	+	-	+	-
2 Minutes	$\bar{X}$	3.22	2.89	3.06	2.61	3.11	2.72	3.17	2.94	2.56	2.67	2.44	2.28
	SD	1.08	0.99	0.71	1.11	0.93	0.87	1.01	0.91	0.89	1.00	0.95	1.35
8 Minutes	$\bar{X}$	3.89	3.33	3.11	3.06	3.56	3.56	3.28	3.44	3.28	2.89	3.44	3.11
	SD	0.31	0.82	0.93	1.03	0.60	0.60	0.81	0.60	0.86	1.15	0.84	0.93

Table III

Comparisons Among Subtest Means

		$\bar{X}_6$	$\bar{X}_5$	$\bar{X}_2$	$\bar{X}_4$	$\bar{X}_3$	$\bar{X}_1$
(4 points) slope--2 units within occasion	$\bar{X}_6 = 5.639$	--	0.55	.278	.778	.833	1.028*
(3 points) slope--2 units within group	$X_5 = 5.694$	--	--	.223	.723	.778	.973*
(2 points) point--2 units within group	$\bar{X}_2 = 5.917$	--	--	--	.500	.555	.750
(2 points) slope--single unit	$\bar{X}_4 = 6.417$	--	--	--	--	.055	.250
(2 points) point--2 units within occasion	$\bar{X}_3 = 6.472$	--	--	--	--	--	.195
(1 point) point--single unit	$\bar{X}_1 = 6.667$	--	--	--	--	--	--

* $p = .01$

Table IV

Multitrait-Multimethod Matrix (Number of Ir..ormational Units as Methods, Point and Slope Information Type as Traits)

		I (single unit)		II (within group)		III (within occasion)	
		pt.	slp.	pt.	slp.	pt.	slp.
I	pt.	(.42)					
	slp.	.07	(.47)				
II	pt.	.93	.80	(.48)			
	slp.	1.00	.62	*1.00	(.37)		
III	pt.	.15	*1.00	.91	.88	(.48)	
	slp.	.15	.76	.20	.21	.31	(.57)

*Note: Actual corrected values greater than 1.