DOCUMENT RESUME

ED 055 839 SE 012 438

AUTHOR	Gueths, J. E.
TITLE	The Interdependence of Instructional Objectives in University General Education Courses.
PUB DATE	Jun 71
NOTE	14p.; Paper Presented at the American Association of Physics Teachers Meeting, June 1971, Beloit, Wisconsin
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*College Science; *Evaluation; General Education; *Measurement; Objectives; *Performance Tests; *Test Construction; Test Reliability

ABSTRACT

The author outlines attempts in measurement in the objective-evaluation component of a descriptive astronomy course for the non-scientist, how the measurement has been made, and some conclusions tentatively drawn as a result of two attempts at these measurements. The learning objectives and summative performance tests have been conceptually included. Four performance tests were developed to measure: (1) observation and inferral skills; (2) analysis; (3) model utilization, numeric and verbal forms, and model synthesis; and (4) synthesis, evaluation, and extrapolation. Only tests 2 and 3 had high enough reliabilities for inclusion in the report. Correlation coefficients are given between the areas of these two tests and with the Henmon-Nelson Tests of Mental Ability. The difficulties of measuring creative or divergent areas and observational and inference skills are recognized. (Author/TS)

THE INTERDEPENDENCE OF INSTRUCTIONAL OBJECTIVES IN UNIVERSITY GENERAL EDUCATION COURSES

J. E. Gueths, Associate Professor of Physics, Wisconsin State University, Oshkosh, Wisconsin

Presented at AAPT meeting, Beloit, Wisconsin
June 17, 1971

Evaluation of student performance is to a physics teacher what experimental measurement is to a physicist. There is no other scientific way to determine the effect of any curricular development, be it a new laboratory exercise or a totally new course, than to measure the effect of that development in terms of student performance; better, to compare terminal performance both before and after the curricular development is implemented.

Unfortunately, the science of measurement of student performance is still quite underdeveloped. We all know the operational definition of a volt, a calorie, or an ampere, but we have not yet really engaged the task of defining student performance in terms that are meaningful to most of us. Therefore, those of us that are interested in the science of science teaching; that is course modeling and measurement of student performance, continue to work with our individual [illegible] systems, each with their individual operational definitions and individual measuring instruments.

This does not mean that the work that is ongoing is valueless. While the individual measurement systems may be unique to each individual science teacher, the conclusions that the individual science teacher draws will be generalizable to the extent that they are correctly defined and described. In the following, we will outline that which we are trying to measure in a university science course serving the non-scientist, how we have attempted

to make the measurement, and finally, some conclusions that we have tentatively drawn as a result of two attempts at these measurements.

The block diagram of the course, a descriptive astronomy course operating in the commonly encountered large lecture, discussion section, and laboratory mode is shown in Figure 1. Most of the papers to be read at this meeting deal with elements of the learning environment. This involves learning strategy, curricular materials, etc. However, we will ignore that part of the course and focus instead on the objective-evaluation component of this university course for non-scientists. The learning objectives and summative performance tests have been conceptually included in the same component. It is a sufficient challenge to deal with those course objectives that are expressed (hopefully) in measurable terms; accordingly we leave instructional objectives that are unmeasureable to the imagination of others. Within this context, learning objectives and performance tests are two inseparable parts of the objective-evaluation component.

The objective system that we are currently using for this course is shown schematically in Figure 2. The desired skills are defined in terms of the process verbs observe, infer, analyze, utilize, synthesize, evaluate, and extrapolate. The composition and relative positioning of the objective areas is similar to the objectives in the cognitive domain of Bloom's taxonomy. As an example of the specification of satisfactory student performance in the Analysis area, the following learning objectives are for the schematic analysis area.

Given a brief written description and associated figures (schematics) depicting a model of a physical system in largely non-mathematical terms, the student will demonstrate the ability to:

a. identify model elements (assumptions, "facts", inferrals, definitions) that are included in the model

b. distinguish between assumptions underlying the model and those elements of the model that follow from direct observation

c. identify the range of validity for the model.

At the moment, the evaluation systems employed in the course primarily consist of multiple choice items. While great care is taken to insure that the student not answer individual process items from memory, this type of examination system makes difficult the measurement of student performance in areas high on the learning hierarchy requiring divergent reasoning. We will revisit this problem shortly.

Four terms are extensively employed in discussing the measurement system. These are; (1) standard score, which is the position of the student with respect to the class mean in terms of the standard deviation of the group performance, (2) the product moment correlation, which is a measure of the interdependency of student performance in two objective areas, (3) reliability, which is a coefficient of internal consistency derived from the correlation of two halves of the same performance test, and (4) the correlation corrected for attenuation, which includes an attempt to normalize a correlation coefficient between two performance areas for their respective non-perfect reliabilities, which, of course, will abnormally depress the correlation coefficient between them.

An alternate method of determining test reliability was employed in this analysis. Because of the relatively small number of items in each area test (usually between 10 and 20) and the associated difficulty in forming two homogenous 1/2 subtests, we employed Kuder-Richardson Equation No. 20, which generates a test reliability from the gross properties of the test and item response distributions. In several cases, we computed test reliabilities both ways, and the differences in the results would not affect any conclusions

we will draw here.

The course is normally populated with 250-300 students. However, during the semester (Spring, 1971) that this curriculum development project was operating, we ran the course under our experimental course number and twenty-nine students enrolled. Of this number, we compiled complete test data on twenty-three, and this data forms the basis of this study.

Four performance tests were administered in the course. The sequence of examination and relative success in measurement is displayed in this overlay (Figure 2). Test I attempted to measure observation and inferral skills. The reliability of the observation test was essentially zero and consequently was excluded from this study. To our knowledge, a reliable convergent measure of observational abilities has not yet been developed. Presumably, this difficulty relates to its being a very basic skill, very difficult to define objectively. Test II focused on analysis, and all area tests exhibited reasonable reliabilities. This test has much in common with conventional physics tests, and the higher test reliabilities are perhaps not surprising in light of the authors experience in regular physics courses. Test III focused on model utilization (in other words, problem sol ing) of both numeric and verbal form and a supposed convergent component of model synthesis. The synthesis sub-test was the most unreliable of these three, presumably reflecting some not surprising difficulty with measuring a creative skill area with a convergent measurement device. Test IV attempted to measure student performance in the areas of synthesis, evaluation, and extrapolation, and reliabilities in all sub-tests were too low to warrant their inclusion in this analysis.

The correlation coefficients between the four areas of the analysis test are displayed in Figure 4. The four areas are:

1. Analysis of data displayed graphically

2. Analysis of data displayed in tabular form and the conversion of tabular data to a graphical form
3. Analysis of a model displayed as a schematic diagram
4. Analysis of a largely non-mathematical scientific argument presented as a written treatise.

These performance areas were found to be reasonably independent of each other with the exception of a strong correlation between the ability to analyze a schematic of a physical system (in this case, the figure conventionally used to treat retrograde motion of a superior planet) and the analysis of tabular data and its conversion to graphical form. One suspects an explanation in terms of mathematical abilities, but we have not been able to document this so far. The strong independence of the schematic and verbal analysis area is also striking (and not totally unexpected) although the rather low reliability involved in both subtests places any conclusion in a tenuous position.

The correlation coefficients between the three areas of the utilization-synthesis test are displayed in Figure 5. These areas attempt to measure student performance in:

1. Using a model (consisting of schematics and data in tabular form) to solve simple numerical problems
2. Using a model (as above) to solve simple problems in verbal form (e.g. which of the following planets would appear to move most rapidly with respect to the stars in the Zodiac as viewed from the surface of the earth?)
3. Ability to identify other changes in a model that will occur as a result of certain changes specified in the test item.

The latter area represents an attempt to measure student performance in the area of synthesis (a decidedly divergent skill area) within a multiple choice

format (a convergent type of examination). We interpret the low reliability of the synthesis sub-test coupled with the strong correlation with both utilization sub-tests as a failure of the synthesis sub-test to perform as we had hoped.

When constructing performance tests that diverge from the commonly encountered knowledge-memory response format, one must be concerned that the measurement instruments not become intelligence tests instead of performance tests. Put another way, our objective is to measure student performance on skill areas that relate to the course, rather than document a pre-existing I.Q.

During the second week of the course, the testing center at WSU-O administered the Henmon-Nelson Tests of Mental Ability, Form A to the group. These are standardized tests of high reliability, and student ability is broken down into the quantitative and verbal areas. To a good approximation, we can assume that the reliabilities of these standardized tests are identically 1.

Figure 6 displays the product moment correlations between student performance on the quantitative and verbal areas of the Henmon-Nelson tests and performance on the 8 area tests developed in this course versus the square root of the reliability of the area test. All the data would fall on the solid line if the area tests were perfectly correlated (when corrected for attenuation) with the Henmon-Nelson tests. We see that the 4 tests involving analysis and two involving utilization are not strongly correlated with I.Q., although the area tests intended to measure inferral and synthesis skills approached being intelligence tests. You will recall that these objective areas lie on both ends of the conditional learning hierarchy.

We have drawn several tentative conclusions based on our efforts in this

area to date.

1. Reliable objective tests based on intellectual skills other than memory can successfully be developed for a large portion of the learning hierarchy we are conditionally using.
2. Reliable objective tests for observational and inference skills are extremely difficult to prepare.
3. If one attempts to measure student performance in regions of the learning hierarchy that lie higher than problem solving, it might be necessary to use non-convergent testing methods. In most cases, divergent testing is difficult to administer in the large lecture courses and this might precipitate some changes in course structure midway through each term.
4. As one approaches testing student performance in either very basic or alternately, the creative areas of the performance objectives, one must be especially careful that the tests do not become intelligence tests rather than performance tests.

FIG. 1. SCHEMATIC OF A CURRICULAR ELEMENT

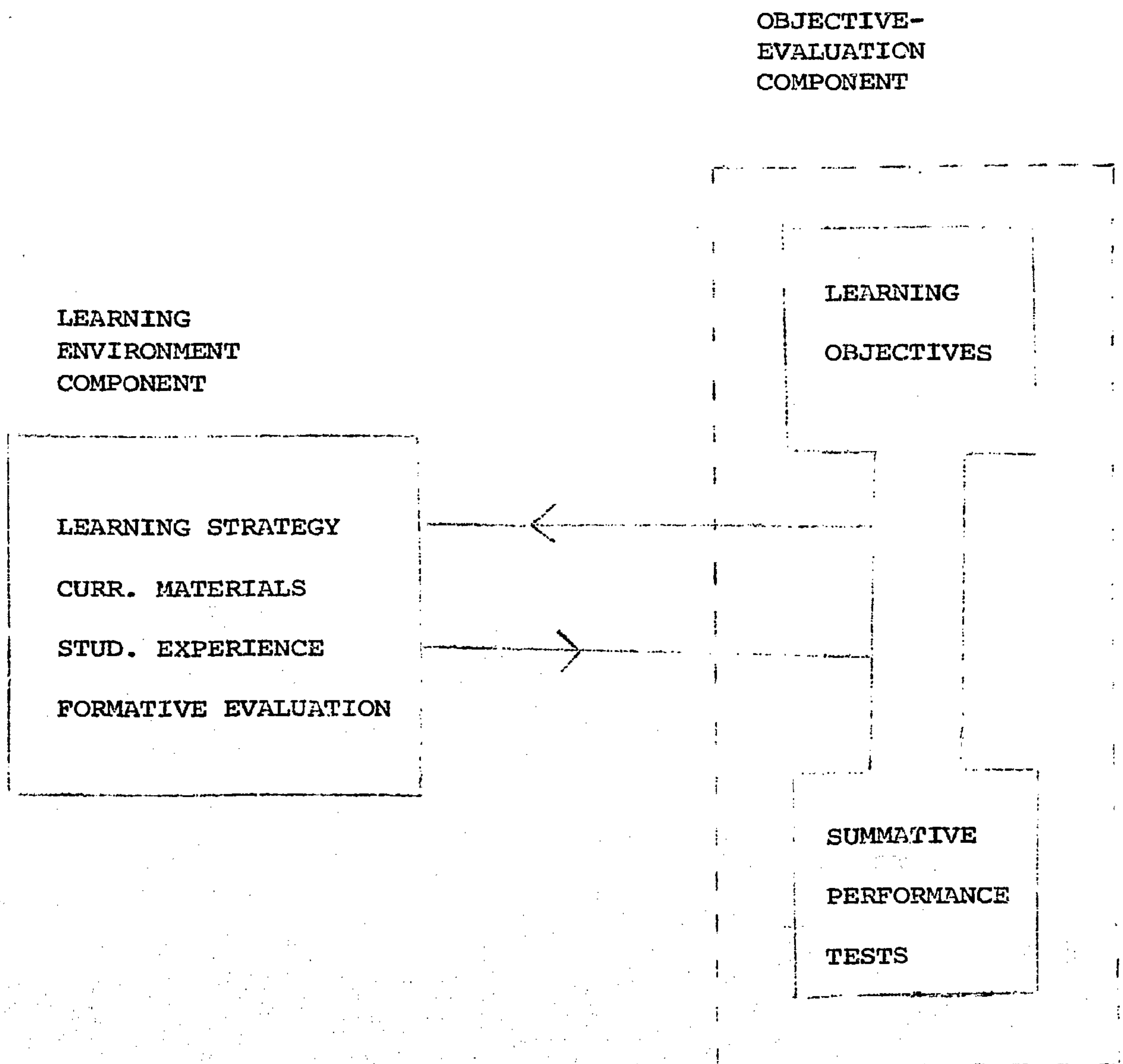

FIG. 2. SCHEMATIC OF OBJECTIVE - EVALUATION SYSTEM

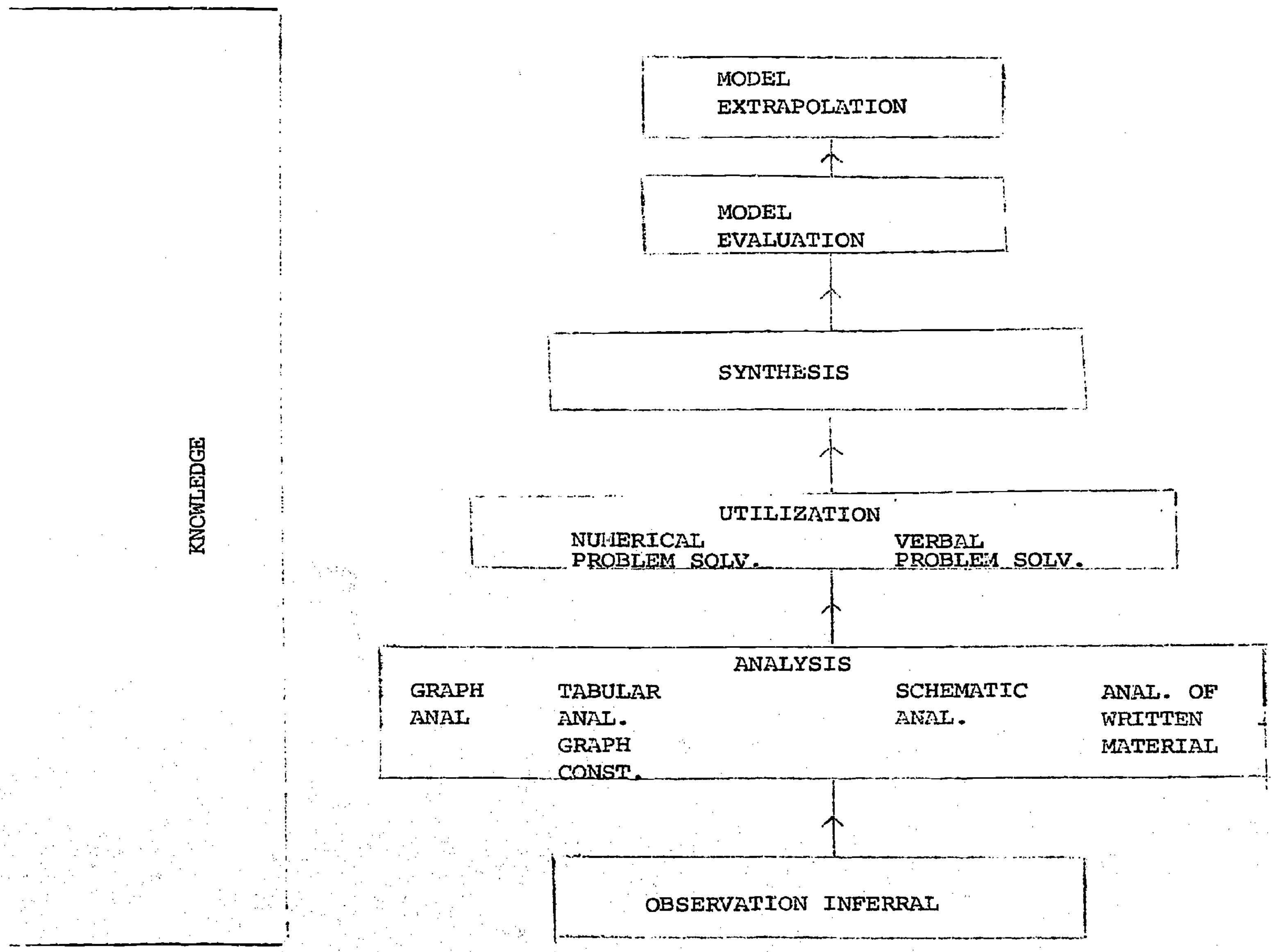

FIG. 2. SCHEMATIC OF OBJECTIVE - EVALUATION SYSTEM

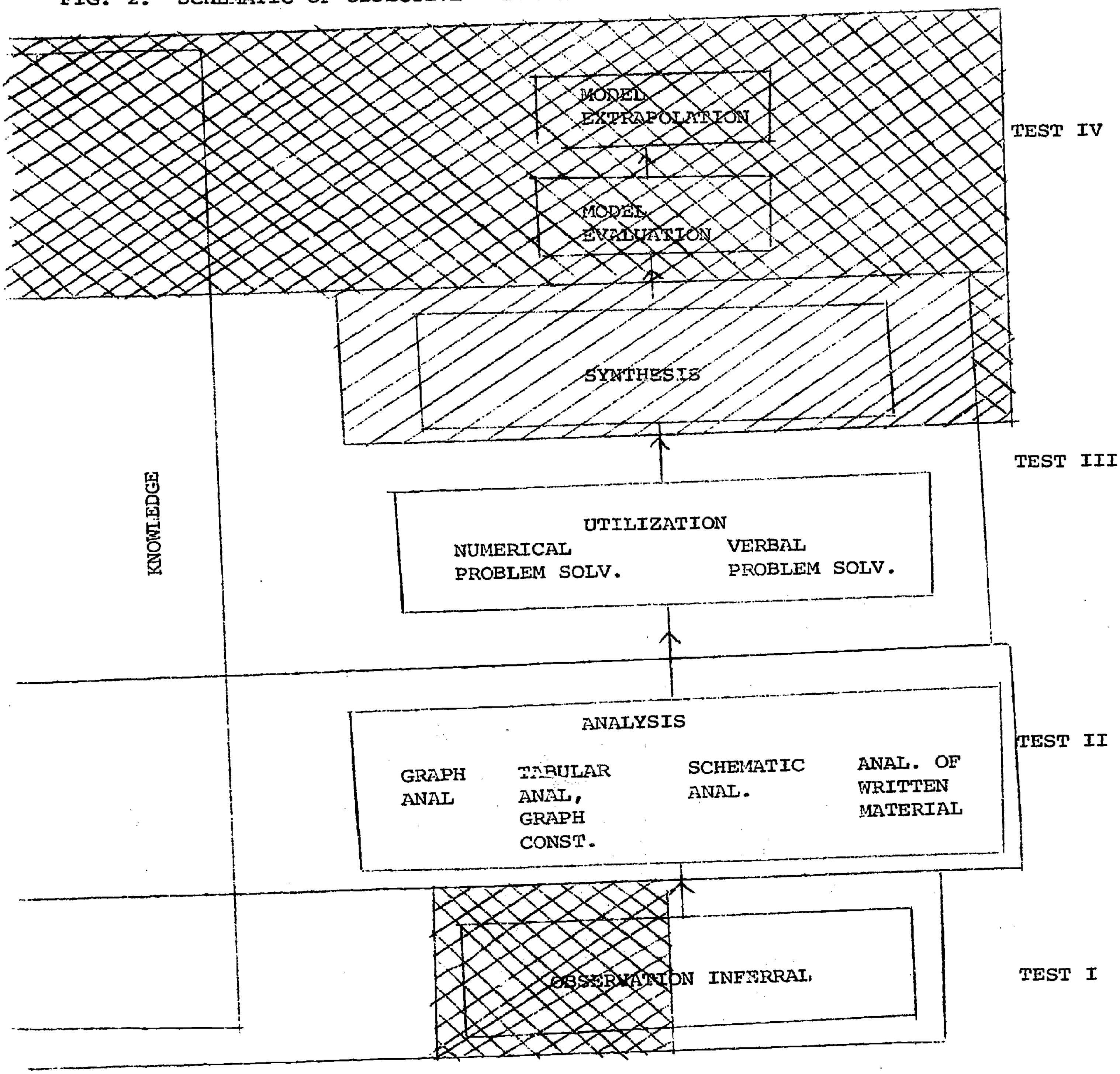

FIG. 3. PARAMETERS OF EVALUATION SYSTEM

STANDARD SCORE

$$s_i = \frac{X_i - \bar{X}_i}{S.D.} \qquad X_i = \#\ \text{CORRECT}$$

product moment correlation

$$r_{ij} = \sum_N \frac{s_i s_j}{N} \qquad N = \#\ \text{STUDENTS}$$

test reliability

$$r_{ii} = \frac{2r\ 1/2}{1 + r\ 1/2}$$

$$r_{1/2} = \sum_N \frac{s_{i\ 1/2}\ s_{i\ 1/2}}{N}$$

or KUDER-RICHARDSON #20

$$r_{ii} = \frac{K}{K-1}\left\{1 - \sum \frac{p(1-p)}{(S.D)^2}\right\}$$

K = # of items

p = fraction of correct responses

S.D. = standard deviation

correlation corrected for attenuation

$$(r_{ij}) = \frac{r_{ij}}{\sqrt{r_{ii} r_{jj}}}$$

FIG. 4

TEST II ANALYSIS TEST R = 0.78

GRAPH ANAL	TAB. ANAL., GRAPH CONST.	SCHEM. ANAL	LIT ANAL	
0.70 (1.00)	0.32 (0.47)	0.31 (0.53)	0.23 (0.46)	GRAPH. ANAL.
	0.65 (1.00)	0.50 (0.86)	0.28 (0.59)	TAB. ANAL, GRAPH CONST.
		0.48 (1.00)	0.11 (0.26)	SCHEMATIC ANAL.
			0.35 (1.00)	LIT. ANAL.

( ) CORRECTED FOR ATTENUATION

FIG. 5

TEST III UTILIZATION - SYNTHESIS R = 0.73

UTIL. NUM.	UTIL. VERB.	SYNTH.	
0.47 (1.00)	0.36 (0.65)	0.44 (1.10)	UTIL. NUM.
	0.66 (1.00)	0.38 (0.81)	UTIL. VERB
		0.34 (1.00)	SYNTHESIS

( ) CORRECTED FOR ATTENUATION

FIG. 6. CORRELATIONS BETWEEN PERFORMANCE TESTS AND QUANTITATIVE AND VERBAL SCORES ON HENMON-NELSON FORM A.