

DOCUMENT RESUME

ED 042 806

TM 000 079

AUTHOR Bart, William M.
TITLE A Construction and Validation of Formal Operational Reasoning Instruments.
PUB DATE Mar 70
NOTE 32p.; Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March, 1970

EDRS PRICE MF-\$0.25 HC-\$1.70
DESCRIPTORS Able Students, *Abstraction Tests, *Abstract Reasoning, Cognitive Measurement, *Cognitive Tests, Content Analysis, Item Analysis, *Logical Thinking, Measurement Instruments, Mental Tests, Psychometrics, Teenagers, Test Construction, Testing, Test Reliability, Test Validity, *Thought Processes

IDENTIFIERS *Piaget

ABSTRACT

In Piaget's developmental psychology the fourth and highest stage of human cognitive development is that of formal operations. The research on formal thought instruments is outlined. This study was designed to construct and validate paper-and-pencil instruments which could be used to select students capable of abstract conceptualization, hypothetico-deductive thought, and combinational reasoning. Three tests in different content areas were developed using item specifications found in the Piagetian literature on formal operations. Items were six-choice logic items with abstruse content. These tests, four Piagetian formal thought tasks, and a measure of verbal intelligence were administered to a sample of above-average teenagers. The formal operational reasoning tests were demonstrated to have substantial content validity, modest concurrent validity, and limited construct validity. Six item structures were found to have uniformly high first principal component factor loadings, validity indices, and reliability indices. The need for further research in this area is pointed out. (DG)

A CONSTRUCTION AND VALIDATION OF
FORMAL OPERATIONAL REASONING INSTRUMENTS

William M. Bart

University of Minnesota

I. Introduction

Piaget has formulated a developmental psychology which indicates an invariant sequence of four qualitatively-distinct stages of human cognitive development. The fourth stage--the stage of formal operations--is characterized by the capacity to consider all the possible relationships in a problematic situation and by the capacity to think in a hypothetico-deductive manner. Formal operations are internalizable, reversible actions which are coordinated in an integrated system and which are based on propositions.

To test for formal operations Piaget and his associates (Inhelder and Piaget, 1958) have formulated a set of experimental tasks which require the application of formal operations for their successful resolution. For example, a billiard game has been used to test for the understanding of the concept of equality of angles of incidence and reflection which, in turn, is a manifestation of the capacity to formulate the binary operation of reciprocal implication.

Formal thought has been measured with the following types of measures: (1) Piagetian tasks; (2) verbal or numerical analogies; (3) test items requiring comprehension of reading passages; (4) logic items. None of these measures have been strictly validated.

In a longitudinal study of four year duration Hughes (1965) tested 40 pupils of average and below average scholastic ability yearly from the age

EDO 42806

TM 000 079

of 11+ years to the age of 14+ years. Four Piagetian tasks including the Equilibrium in the Balance task were used. The task scores on the fourth testing were correlated with other test scores such as those of numerical analogies and non-verbal intelligence test. With a principal component analysis all of the tasks were found to have high correlations (.57 - .81) with the first principal component.

Lovell and Butterworth (1966) tested 60 pupils with an array of measures testing for the schema of proportion including the Equilibrium in the Balance task and the Projection of Shadows task. From a principal component analysis of the scores they found that all of the measures correlated highly with the first principal component. Using an array of formal operations tasks and tests Lovell and Shields (1967) tested 50 pupils ranging in age from 8 to 10 years and having verbal IQ's in excess of 140 as measured by the Wechsler Intelligence Scale for Children. Using a principal component analysis they found that the tasks have quite high correlations with the first principal component including the Equilibrium in the Balance task and the Colorless Chemicals task with .83 and .72 first component correlations respectively. From these three studies which indicate that Piagetian formal thought tasks have a high first principal correlations one can infer that the tasks have substantial concurrent validity.

Research on formal operations using verbal or numerical analogies include a study by Lovell and Butterworth (1966) heretofore cited. English researcher Lunzer (1965) has argued that both verbal and numerical analogies require the application of formal operational skills for second-order relations need to be recognized for the solution of the analogy items; the

capacity to formulate second-order relations is a characteristic of formal thought. Lovell and Butterworth (1966) employed 20 tests (e.g., verbal analogies) involving proportion and showed that a central intellectual ability underlies all of these tests; the capacity to understand proportions relates to the schema of proportion which is an aspect of formal thought.

Research on formal operations using analogy items, though sparse, has given some credibility to the statement that analogy items are valid measures of formal thought. Analogy items have content validity for they require the recognition of second-order relations and the use of the schema of proportion and analogy tests have concurrent validity for they have high positive correlations with other measures of formal thought (e.g., tasks). However, more research needs to be done on the validation of these measures for there is substantial query as to whether analogies test for a broad enough range of behaviors proper to the stage of formal operations.

Studies on formal operations using test items requiring comprehension of reading passages have been done by a variety of researchers (e.g., Stone, 1966). Case and Collinson (1962), Goldman (1965), and Hallam (1967) have employed reading passages in such areas as literature, religion and history. The subjects in those three studies were instructed to read the passages and then to answer a few questions. The oral responses were recorded and then scored with the use of protocols indicating the qualities of responses proper to each of the three highest Piagetian cognitive stages.

Mary Ann Stone (1966) used a set of three reading passages in literature, social studies, and science respectively and with a forty-item multiple-choice test on each passage which demand either recall or application

skills for each item. She contended that comprehension and application behaviors as discussed by Bloom, et al (1956) are proper to formal thought, whereas recall behaviors are proper to lower-stage thought. She determined that competency at application items is higher and more homogeneous across content areas for older pupils than for younger pupils. Though her contention that tests used were valid measures of formal thought is highly questionable since no form of validity was firmly established, she did demonstrate that competency in thinking in various content areas (i.e., horizontal décalage) increases with age.

Research on formal operational thought using logic items has been sparsely done (e.g., Morf, 1957). Albert Morf (1957) who is an associate of Jean Piaget at Geneva stated that any problem that demands an individual reason deductively from a set of hypothetical premises with unary and binary connectives (e.g., if . . . then) is a formal operational problem. Shirley Ann Hill (1960) used logic items testing for the sentential logic, the classical syllogism, and the logic of quantification. For each item the subject was asked to distinguish between a necessary conclusion and the negation of a necessary conclusion. Hill contended that these items tested for hypothetico-deductive reasoning which is a crucial characteristic of formal thought.

However, O'Brien and Shapiro (1968) contended that her items were not content valid for her items did not, in addition, demand that the pupil test the logical necessity of a conclusion. They determined that though youngsters between 6 and 8 years of age are able to discriminate between a necessary conclusion and its negation they are unable to test the logical necessity of a conclusion. Thus they concluded that hypothetico-deductive

reasoning ability cannot at all be attributed to young children. Logic items as formal thought measures have so far manifested only modest content validity.

In general, there presently exist no formal thought instruments that have been extensively validated. The research reported here deals with a construction and validation of formal operational reasoning instruments.

II. Plan of the Study

A. Subjects

A sample of ninety adolescents from Chicago area schools was used in this study. Thirty scholastically above-average students were selected from each of the three age levels: 13 years of age, 16 years of age, and 19 years of age. Since the stage of formal operations was examined, it was assumed that most pupils over 13 years of age inclusive and with above-average scholastic achievement would have formal operational capabilities according to the Piagetian finding that formal operations develop in Swiss children during the 12-15 year period of life (Inhelder and Piaget, 1958). There were 32 males and 58 females in the sample. This condition of disproportionate sex sampling should not detract from the results of the study for evidence has accumulated that there are no sex effects with respect to formal operational skills (e.g., Stone, 1966; O'Brien and Shapiro, 1968).

B. Instruments

1. Piagetian Tasks.

To test for attainment to the stage of formal operations four formal operations tasks devised by Piaget and his associates were employed. The specifications and testing procedures required for these tasks have been elaborated by Inhelder and Piaget (1958), by Lovell (1961), and by Hughes

(1965); the specifications and testing procedures complied with in this study.

The following four tasks have been determined by Piaget and his associates to test for formal operational ability:

- (a) the Oscillation of a Pendulum task which tests for the operations of exclusion;
- (b) the Conservation of Motion on a Horizontal Plane task which tests for the conservation of motion concept;
- (c) the Equilibrium in the Balance task which tests for the understanding of the physical principle of a balance;
- (d) the Projection of Shadows task which tests for understanding of the physical principle relating the size of a shadow to the size of an object projected and to the distances of the object from the light source and from the surface of the shadow.

The plane and pendulum tasks require the experimental manipulation of variables to confirm certain hypotheses. The balance and shadow tasks require the discovery of such relations as proportionality and reciprocity in physical systems.

These four tasks were administered to each subject in the sample and the response to each task was given a rating of one of the stage levels (I, II-A, and II-B, and III-A and III-B) used by Piaget and his associates to grade these tasks. The protocols used in this study were strictly complied with in giving the ratings. A rating of III-A or III-B was given to concrete operational responses; a rating of I was given to preoperational responses.

2. Formal Operational Reasoning Instruments.

These instruments are measures of formal operational thought. Each adolescent was administered three of these instruments: three formal operational reasoning instruments set in the content areas of biology, literature, and history respectively.

Piaget (1963) contends that the ability to accept absurd premises (e.g., there was a dog with six heads) as such and to reason from these premises in a purely deductive manner is formal operational. Morf (1957) states that the ability to reason deductively from a set of premises in which unary and binary connectives (e.g., if . . . then) is also formal operational. The following item is an example of formal operational reasoning devised by Morf (1957):

I think of an animal. If the animal has long ears, it may be either an ass or a mule. If my animal has a big tail, it is either a mule or a horse. Now, I want an animal with both long ears and a big tail. What can it be?

Thus, any verbal or written item that requires an acceptance of a set of absurd premises in which unary and binary connectives are used and that requires a deduction problem to be solved based on the absurd premises is a valid test of formal operational thought according to the considerations of Piaget and of Morf. This general specification for the formal operational reasoning instruments was adhered to in the construction of many items. The following specifications were complied with in the item construction:

- i. Each item has either absurd (contrary-to-fact) declarative premises or imaginary declarative premises--i.e., each premise in each item must be

either contrary to a fact that the subject knows to be a fact or must be an imaginary statement which has no concrete referents and is beyond the experience of the subject. An example of a contrary-to-fact premise is the following: William Shakespeare wrote Tom Sawyer, but he did not write Hamlet. An example of an imaginary premise is the following: Unicorns travel only in pairs. The specification referring to absurd premises is attributable to Jean Piaget (1963) and the specification referring to imaginary premises is derivable from the discussion on formal thought by Flavell (1963).

ii. Each item has unary and binary connectives (e.g., "if . . . then," "but", "and", "not", "neither . . . nor") being used in the premises. An example of such a premise in which the binary connective "but" will be used is the following: William Shakespear wrote Tom Sawyer, but he did not write Hamlet. This specification is attributable to Albert Korf (1957) who is an associate of Piaget.

iii. The task for each item requires a simple deduction through the use of either propositional rules of inference or quantificational rules of inference in order for the validly deducible response to be recognized. An example of such an item with the content area being biology and the primary rule of inference to be used being modus tollens is the following:

A. If butterflies can swim, then butterflies have gills.

Butterflies do not have gills, but they have fins. Therefore . . .

- a. All butterflies can swim.
- b. Either butterflies swim or they have gills.
- c. If butterflies fly, then they swim.
- d. Butterflies cannot swim.
- e. Some butterflies have no fins.
- f. Butterflies swim but have no fins.

The correct response to this item is d.

This specification is attributable to both Piaget (1963) and Morf (1957).

iv. It is indicated to the student in each test that he is to assume that the premises for a given item are true. This specification is attributable to Piaget (1963).

v. Each item in the tests is of the multiple-choice type with six choices.

Items were constructed so that each required thought processes that are appropriate to the stage of formal operations but not appropriate to the stage of concrete operations--i.e., each item to be used must comply to specifications i-iv for formal operational reasoning tests that have heretofore been stated and that are attributable to Morf and Piaget. Two high school teachers were chosen and trained in the appraisal of items as conforming to the specifications cited. If they both agreed that the items chosen complied to the specifications cited, then content validation of the items will have been achieved. There was no disagreement between the two raters for they both agreed that the items complied to the specifications cited; thus, a content validation of the items was achieved.

The observation that there was no disagreement between the raters as to the compliance of the items to the specifications may seem extraordinary. However, upon a closer examination of the procedure used for content validation this observation may appear to be more reasonable. First of all, both high school teachers who were raters had some familiarity with symbolic logic to the extent that they knew basic logical rules of inference and knew the English notation used in symbolizing certain logical statements (e.g., the statement "John went to the store and Mary went home" could be symbolized as "p.q." where p refers to "John went to the store", q refers to "Mary

went home", and "." refers to "and"). Their familiarity with symbolic logic proved to be an aid in their scrutiny of the test items.

The items were constructed and a statement of the specifications were provided to each rater. Each rater was asked to examine each item and to identify any item not complying to the item specifications. First of all, each item was found to have either absurd or imaginary premises by each rater. Secondly, each item was examined and found to have unary and/or binary connectives (e.g., "not", "either . . . or", "and") being used in the item by each rater. Thirdly, each rater determined that each item required some logical rule of inference to be used for its correct resolution; in this phase the teachers used their familiarity with symbolic logic to ascertain item compliance to specification iii. Lastly, each rater indicated that the direction for the items made it sufficiently clear that the premises in each item were to be assumed to be true by each subject. Due partly to the prior training in symbolic logic of the raters and due to the pointedness, simplicity, and clarity of the item specifications the raters were able to scrutinize the items for possible compliance to the item specifications and determined that the items had content validity.

After a set of content valid items were constructed, a pre-testing of the items was enacted to determine those items with high-point biserial correlations, high factor loadings with the first principal component, and item difficulties in the .10 - .90 range. A subset of about 30 items was selected from those items to form the basic items in the instrument. Versions of the test were constructed maintaining the formal logical structures of the items but set in the three content areas--biology, history, and literature. The testing time for each of the tests was about forty

minutes.

The formal logical structures of the thirty items chosen were set in terms of the Polish notation of symbolic logic developed by J. Lukaseiwicz, a noted Polish logician. These thirty items were set in various contents in the pre-test of sixty items. It may be noted that the correct response for any given item forms the only consistent and valid formula with the premises given. The other five responses for any given item relate to inconsistent formulae that are invalid for some choice of truth values for the constituent premises determines true composite premises and false conclusion responses. After the thirty items were selected from the pre-test, the items were randomly ordered to form the standard ordering of items in the formal operational reasoning instruments.

The measure of competency with formal operations is the mean score (mean number of items answered correctly) for the three reasoning scores; this score is the formal operational competency score. An individual is adjudged to be capable of formal operational thought if he obtains a formal operational competency score greater than the upper 95% confidence limit for the guessing score ($n/6 + \sqrt{5n}/3$ where n is the number of items in the test--Gulliksen, 1950) for the three reasoning scores for that individual.

3. Measure of General Intellectual Ability.

A measure of verbal intelligence entitled the Experimental Omnibus Vocabulary Test developed by Frederick Davis is the measure of general intellectual ability used in this study. The forty items used in this test were selected from a larger sample of vocabulary items as conforming

to a unidimensional model of verbal intelligence. In part, the test scores were used in the construct validation procedure for the three formal operational reasoning instruments.

The three formal reasoning tests and the vocabulary tests were administered to the ninety adolescents in group-testing settings and the four Piagetian tasks were administered to each adolescent individually.

III. Results

A test has content validity if the items in the test require behaviors for their successful resolution that are proper to the trait being measured (Cronbach, 1960). The three formal reasoning tests were found to have content validity. For example, the reasoning tests were found to have all of their items fulfilling the Genevan specifications i-iv for formal operational reasoning tests heretofore cited.

A test has concurrent validity if the test correlates highly positively with direct tests measuring the same trait as the initial test (Cronbach, 1960). Concurrent validity of the formal reasoning tests was to be determined in two phases. The first phase entails the examination of the correlations between the formal operational task scores and each of the sets of formal reasoning scores and the total reasoning scores: Table 1 depicts these correlations.

Table 2 indicates the lower bounds of the correlations corrected for attenuation between the total task scores and the four formal reasoning scores cited in Table 1. The correlations in Table 2 are lower bounds for in the computation of these correlations it was assumed that the reliability of the composite task was 1.00. Thus the coefficient indicating the relation

between the set of four tasks and a given formal reasoning test conditional on the usage of perfectly reliable instruments is at least the corresponding correlation designated in Table 2.

From an examination of the correlations in Tables 1-2 it can be stated that the relation between the formal reasoning scores and the total task scores for four Piagetian tasks is moderate even if perfectly reliable instruments are used.

Being not uniformly high the correlations in Table 1 lend weight to the contention that the formal reasoning scores are moderately related to the total task score for the four Piagetian tasks used. Thus the first phase of the concurrent validation has provided information attesting to the modest concurrent validity of the separate formal reasoning tests. However, when these tests are combined, the concurrent validity (with or without attenuation) is relatively high.

The second phase of concurrent validation entails the examination of a contingency table relating the placement of individuals into cognitive stages according to their formal operational task scores to the placement of individuals into cognitive stages according to formal operational competency scores. No adolescent subject was found to be at the pre-operational stage of thought according to their Piagetian task performances. The formal reasoning tests can only specify formal thought capabilities from non-formal thought capabilities. All adolescent subjects were placed at either the concrete stage of thought of the formal stage of thought.

Individual task scores greater than three were judged to be formal operational and thus formal operational task scores greater than 12 were

classified as formal operational. Subjects with task scores in the 5-12 range were classified as concrete operational. The upper 95% confidence limit for the guessing score for the three reasoning tests is 9.1 which equals $n/6 + \sqrt{5n}/3$ where $n=30$ (Gulliksen, 1950). Subjects receiving competency scores greater than 9.1 were classified as concrete operational. Adolescent subjects receiving competency scores less than or equal to 9.1 were classified as concrete operational. Table 3 indicates the placement of the 90 subjects into the two highest cognitive steps according to the Piagetian tasks and according to the formal reasoning tests.

As can be seen in Table 3, 86 subjects were found to be at the stage of formal operations according to the two sets of measures. Also 95.5% of the adolescent subjects were adjudged to be at the stage of formal operations by both methods of stage measurement. The capacity of the set of four Piagetian tasks to measure subjects at the stage of formal operations is to a great extent shared by the set of formal reasoning tests. However, this phase of the concurrent validation remains inconclusive for the alleged capacity of the formal reasoning tests to distinguish subjects at the stage of formal operations was not verified in this study. Hopefully, in the future a wide variety of subjects could be chosen and tested with the tasks and reasoning tasks and then could be classified into cognitive steps according to their respective sets of responses. The classifications according to the task scores and according to their respective sets of responses. The classifications according to the task scores and according to the test scores could be examined and compared with the use of a contingency table and then the discriminative quality of the formal

reasoning tests could be determined. In this manner the phase of the concurrent validation of the formal reasoning tests could be accomplished.

It may be noted that no capacity of the formal reasoning tests to classify subjects into lower cognitive stages has been acknowledged, thus the classificatory range of the Piagetian tasks is recognized as being greater than the classificatory range of the formal reasoning tests.

Construct validity of the formal reasoning tests was determined in three phases. The first phase involved some of the techniques of convergent and discriminant validation proposed by Campbell and Fiske (1959) on the examination of a multitrait-multimethod matrix. Convergent validity of a set of test measuring a given trait is demonstrated if the tests have high positive correlations with other tests employing a different method measuring the same trait. It was hypothesized that the correlations between the formal reasoning tests and the Piagetian tasks will be large and positive thus indicating the convergent validity of the formal reasoning tests. Discriminant validity of a set of tests measuring a given trait is demonstrated if the test have small correlations with measures of a different trait but employing a similar method. It was thus hypothesized that the correlations between the three formal reasoning tests and the measure of verbal intelligence will be quite low, thus indicating the discriminant validity of the formal reasoning tests.

Tables 4-7 indicate the intercorrelations among the eight constituent cognitive variables used in this study for each of the three age levels and the total sample. Most of the correlations in the rectangular sub-matrices with the dotted lines are modestly significant and positive, thus attesting to the limited convergent validity of the formal reasoning tests; the formal

reasoning tests are, in general, moderately correlated with the Piagetian tasks.

The rectangular sub-matrices with the solid lines in Tables 4-7 relate to the discriminant validation aspect of this phase for they indicate the correlations between the measure of verbal intelligence and the three constituent measures of formal thought. Most of these values are modestly significant. In addition, these correlations are, in general, of the same magnitude as the correlations indicating convergent validity of the formal reasoning tests. These conditions indicate that the formal reasoning tests have little if no discriminant validity for two reasons: (1) the formal reasoning tests have modest, not small as hypothesized, correlations with a measure of a different trait (i.e., verbal intelligence) as they are with measures of the same trait (i.e., formal thought).

The second phase of the construct validation entailed a scrutiny of the age level means and standard deviations for the vocabulary test scores and the formal reasoning test scores. Table 8 indicates this information. It was hypothesized that the means of the vocabulary test scores will indicate a decided positive monotone trend, whereas the formal reasoning test score means increase from the 13 year age level to the 16 year age level but then level off and show no significant increase from the 16 year age level to the 19 year age level. That hypothesis is derivable from the observations that verbal intelligence (e.g., vocabulary size) continues to grow well into adulthood (Guilford, 1967) but that formal reasoning becomes well established in a relatively short period of time after its emergence (Inhelder and Piaget, 1958).

As can be discerned in Table 8 there is a definite positive monotone trend among the vocabulary test means, whereas there is no significant increase for any of the formal reasoning test measures between the 16 year age level and the 19 year age level. The statistics in Table 8 indicate that as age increases the vocabulary size tends to increase and becomes more varied, thus substantiating an aspect of the hypothesis being considered. Also, statistics in Table 8 indicate that between 16 and 19 years of age formal operational skills become somewhat fixed. These two trends in the scores discernible in Table 8 attest to the hypothesis that formal operations and verbal intelligence comply to two different growth patterns.

The third phase involved the examination of some of the correlations and partial correlations among the total task scores, the formal operational competency scores, and the vocabulary scores. It was hypothesized that verbal intelligence is not the primary component in the formal operational relationship between the Piagetian tasks and the formal reasoning tests. This hypothesis was confirmed if the partial correlation between the formal operational competency scores holding vocabulary scores constant was similar to the correlation between the formal operational task scores and the formal operational competency scores.

A second aspect of this third phase designated the hypothesis that little remains of the relationship between verbal intelligence and formal reasoning when the Piagetian formal task component is removed. This was confirmed if the partial correlation between the vocabulary scores and the formal operational competency scores holding formal operational task scores constant was appreciably less than the correlation between the vocabulary scores and the

formal operational competency scores. Table 9 indicates the four correlations relevant to this phase of the construct validation.

From the correlations cited in Table 9 it can be determined that the measures of verbal intelligence accounts for only 27.3% of the variance shared by the Piagetian tasks and the formal reasoning tests and that the two correlations to be considered in the first aspect of the third phase are quite similar as hypothesized. Also, the partial correlation between the vocabulary scores and the formal operational competency scores holding formal operational task scores constant (.159) is appreciably less than the correlation between the vocabulary scores and the formal operational competency scores (.358) as hypothesized. Thus this phase of the construct validation of the formal reasoning tests provides evidence attesting to the construct validity of the formal reasoning tests.

To summarize the validation procedure findings, the formal operational reasoning tests were demonstrated to have substantial content validity, modest concurrent validity, and limited construct validity.

From an examination of certain statistical and psychometric properties of the tests used in the study certain findings on the item structures can be stated. Five item structures had relatively high average validity indices (in excess of .100), high average reliability indices (in excess of .130), relatively high first principal component factor loadings, and moderately high item difficulties (in the .500 - .760 range). The high average validity indices of these item structures, for example, indicated that the performance on any item with any of these item structures is highly related either with performance on the four Piagetian tasks used or with the formal operational

competency score. No discernible pattern was evident among these item structures for four of the item structures were drawn from the quantificational logic and one item structure was drawn from the logic of relations.

More research is needed to determine those logical components (e.g., presence of connective "not") in a formal reasoning item with an item structure responsible for the validity and difficulty of the item. Also research is needed to determine whether items with more abstract content or with more complex constituent sentences are more difficult than other items set in different contents but with the same formal logical structure.

In general, highly reliable formal reasoning tests with items having high reliability and validity indices form an objective for formal reasoning test construction and would provide more accurate and valid measures of formal thought. Also formal reasoning tests with items complying with a factorial design with types of logical components (e.g., binary connectives such as "and") designating the factors used could be used to determine those qualities of the items that would influence item discrimination and item difficulty.

IV. Discussion

The attempt to construct and validate paper-and-pencil formal operations tests was somewhat successful. The formal reasoning tests developed in this study could be used to determine the level of formal cognitive functioning for each adolescent in a school. However, valid, "pure" paper-and-pencil measures of formal thought were not developed; more research is needed to resolve the methodological problems in the construction of such a valid, "pure" instrument. However, the set of procedures employed in this study

would provide a reasonable basis from which such instrument development could take place. Valid, "pure" formal thought tests would have considerable value for not only the practical purposes of measuring level of cognitive development but also for instituting fertile psychological research in formal thought and adult cognition.

In general, there still presently exist no strictly validated instruments testing for formal operations. With validated formal operations tests easy to administer, educators could determine those students capable of higher cognitive functioning. In addition, a standardized developmental scale of reasoning battery consisting of paper-and-pencil instruments that measure formal thought capabilities and other paper-and-pencil instruments that test for other Piagetian stage behavior patterns could be used extensively and inexpensively to determine the level of cognitive development of each member of the school population and to diagnose the cognitive inabilities of the mentally retarded. Thus research on the measurement of operational thinking has not only a theoretical relevance but also extensive practical ramifications.

It is quite possible that the evaluation of the subject matter achievement and the measurement of the cognitive development will be unified through test construction from a Piagetian framework. For example, tests demanding the same set of operational skills may be set in various content areas to test for the generalizability of the operational skills and to test for achievement in the content areas. Thus psychologically-parallel achievement tests could be constructed that would indicate the level of cognitive development and the achievement of an individual for a set of content areas. It is anticipated that these contentions may have considerable effect on measurement

and evaluation practices in schools.

Table 1

Product-Moment Correlation Coefficients Between Formal Operational Task Scores
and Each of the Three Sets of Formal Operational Reasoning Scores
and the Total Formal Reasoning Scores

Piagetian Tasks	Formal Operational Reasoning Tests			Total Score
	Biology	History	Literature	
13 year old	.220	.096	.404	.333
16 year old	.513	.593	.523	.624
19 year old	.477	.484	.515	.571

Table 2

Lower-bounds for Pearson Product-Moment Correlation Coefficients
Corrected for Attenuation Between Formal Operational Task Scores
and Each of the Three Sets of Formal Operational Reasoning Scores
and the Total Formal Reasoning Scores

Piagetian Tasks	Formal Operational Reasoning Tests			Total Score	
	Age-level	Biology	History		Literature
Total Group		.566	.535	.630	.792

Table 3

Contingency Table Indicating the Number of Subjects at the Concrete Stage of Thought and at the Formal Stage of Thought According to the Task Scores and According to the Competency Scores

		Formal Reasoning Tests		
		Concrete	Formal	Totals
Piagetian Tasks	Concrete	0	4	4
	Formal	0	86	86
	Totals	0	90	90

Table 4

Lower-Triangular Correlation Matrix for the Seven Measures of Formal Thought
and the Measure of Verbal Intelligence for the 13 Year Old Group

Variable	1	2	3	4	5	6	7	8
1. Shadows Task								
2. Balance Task	.746							
3. Pendulum Task	.379	.566						
4. Conservation Task	.368	.609	.377					
5. Biology Test	.342	.249	-.103	.191				
6. History Test	.209	.070	-.052	.069	.562			
7. Literature Test	.253	.474	.286	.261	.504	.383		
8. Vocabulary Test	.133	.278	-.086	.072	.101	-.004	.384	

Table 5

Lower-Triangular Correlation Matrix for the Seven Measures of Formal Thought
and the Measure of Verbal Intelligence for the 16 Year Old Group

Variable	1	2	3	4	5	6	7	8
1. Shadows Task								
2. Balance Task	.817							
3. Pendulum Task	.682	.609						
4. Conservation Task	.552	.596	.528					
5. Biology Test	.330	.458	.500	.503				
6. History Test	.524	.538	.533	.415	.626			
7. Literature Test	.460	.499	.393	.429	.641	.582		
8. Vocabulary Test	.316	.428	.425	.551	.202	.130	-.019	

Table 6

Lower-Triangular Correlation Matrix for the Seven Measures of Formal Thought
and the Measure of Verbal Intelligence for the 19 Year Old Group

Variable	1	2	3	4	5	6	7	8
1. Shadows Task								
2. Balance Task	.528							
3. Pendulum Task	.530	.487						
4. Conservation Task	.638	.325	.261					
5. Biology Test	.315	.596	.337	.229				
6. History Test	.448	.552	.122	.330	.460			
7. Literature Test	.390	.558	.319	.318	.699	.664		
8. Vocabulary Test	.228	.431	.196	.171	.407	.248	.426	

Table 7

Lower-Triangular Correlation Matrix for the Seven Measures of Formal Thought

and the Measure of Verbal Intelligence for the Total Sample

Variable	1	2	3	4	5	6	7	8
1. Shadows Task								
2. Balance Task	.731							
3. Pendulum Task	.560	.594						
4. Conservation Task	.511	.538	.448					
5. Biology Test	.379	.460	.328	.327				
6. History Test	.396	.388	.221	.269	.508			
7. Literature Test	.416	.536	.458	.375	.654	.461		
8. Vocabulary Test	.260	.405	.376	.357	.325	.134	.394	

Table 8

Statistics of Five Cognitive Test Scores Over Three Age-Levels

Age Level	Statistic	Experimental	Biology	History	Literature	Mean
		Omnibus Vocabulary Test Score	Formal Thought Test Score	Formal Thought Test Score	Formal Thought Test Score	Formal Operational Competency Score
13 Year Old	mean	15.733	17.366	18.466	15.000	16.940
	s.d.	4.016	3.633	2.661	4.961	3.059
16 Year Old	mean	18.100	20.200	18.366	19.733	19.440
	s.d.	4.830	3.325	3.652	3.694	3.068
19 Year Old	mean	24.166	19.933	18.666	19.766	19.460
	s.d.	5.337	4.016	3.325	3.710	3.166
Total Sample	mean	19.333	19.166	18.500	18.166	18.613
	s.d.	5.907	3.848	3.205	4.693	3.286

Table 9

Some Correlations Among Cognitive Measures for Total Sample

1. Correlation between formal operational competency scores and formal operational task scores	= .564
2. Partial correlation between formal operational competency scores and formal operational task scores holding vocabulary scores constant	= .488
3. Correlation between formal operational competency scores and vocabulary scores	= .358
4. Partial correlation between formal operational competency scores and vocabulary scores holding formal operational task scores constant	= .159

References

- Case, R. D. & Collinson, J. M. The development of formal thinking in verbal comprehension. British Journal of Psychology, 1962, 32, 103-111.
- Cronbach, L. J. Essentials of psychological testing. New York: Harper and Row, 1960.
- Goldman, R. J. The application of Piaget's schema of operational thinking to religious story data by means of the Buttnan scalogram. British Journal of Psychology, 1965, 35, 158-70.
- Hallam, R. N. Logical thinking in history. Educational Review, 1967, 19, 183-202.
- Hill, S. A. A study of logical abilities in children. Unpublished doctoral dissertation, Stanford University, 1960.
- Hughes, M. M. A four year longitudinal study of the growth of logical thinking in a group of secondary modern school long. Unpublished M.Ed. thesis, University of Leeds, 1965.
- Inhelder, B & Piaget, J. The growth of logical thinking from childhood to adolescence. New York: Basic Books, Inc., 1958.
- Lovell, K. A. & Battenworth, I. B. Abilities underlying the understanding of proportionality. Mathematics Teaching, 1966, 37, 5-9.
- Lovell, K. A. & Shields, J. B. Some aspects of the study of the gifted child. British Journal of Educational Psychology, 1967, 37, 201-208.
- Lunzer, E. A. Problems of formal reasoning in test situation. In Mussen, P.H. (Ed.), Europe n research in cognitive development. Monographs of the Society for Research in Child Development, 1965, 30, 19-46.

Morf, A. Les relations entre la logique et le langage lors du passage du raisonnement concret au raisonnement formel. In Apostol, L., Mandelbrot, B., & Morf, A. Logique, langage et theorie de l'information. Etudes d'epistemologie genetique. Vol. 3. Paris: Presses Univer. France, 1957, 173-204.

O'Brien, T. C. & Shapiro, B. J. The development of logical thinking in children. American Educational Research Journal, 1968, 5, 531-542.

Stone M. A. The development of the intersituational generality of formal thought. Unpublished doctoral disseration, University of Illinois, 1966.