# The Hierarchical Structure of Formal Operational Tasks

**William M. Bart**
**University of Minnesota**

**Donna M. Mertens**
**University of Kentucky**

The hierarchical structure of the formal operational period of Piaget's theory of cognitive development was explored through the application of ordering theoretic methods to a set of data that systematically tapped the various formal operational schemes. The results suggest that the tasks within some schemes are empirically equivalent. While the response patterns were quite varied, the results do suggest that some common structure may underlie performance on the tasks, thus supporting Piaget's notion of the integrative structure of the period.

Inhelder and Piaget (1958) proposed that the most advanced period of cognitive development is termed the period of formal operations, in which formal operational schemes (or operational skills) emerge concurrently and become coordinated into a structure; those schemes are necessary for an individual to solve problems in a formal operational manner. Empirical research with regard to the integrative structure of formal operations is necessary to understand this level of Piaget's theory of cognitive development.

Three problems emerge from previous research into the nature of this integrative structure. First, research which focuses on intraindividual differences is scarce (Martorano, 1977). Most formal-level research has investigated group performance differences. Second, studies which have focused on intraindividual performance differences have not used a method of analysis that allowed exploration of the relationships among the tasks and, thus, of the integrative nature of the stage (Jackson, 1965; Lovell, 1961; Martorano, 1977). Third, studies which have used an analysis method designed to investigate hierarchical relationships have used only a limited subset of tasks and have exhibited no systematic effort to use tasks that tap the same operational schemes (Bart & Airasian, 1974).

One purpose of the present study was to investigate the hierarchical structure of formal operations through the application of ordering theoretic methods to a set of data that systematically tapped the various formal operational schemes. This would provide a more comprehensive test of the nature of the schemes and the tasks designed to measure them and would give insight into the integrative structure of the cognitive period.

## Method

The data that were analyzed in the present study were collected by Martorano (1977). She administered 10 of the formal operations tasks to 4 groups of 20 6th-, 8th-, 10th-, and 12th-grade adolescents. The tasks were representative

of several of the formal operations schemes and, with the exception of the correlations scheme and the colored tokens scheme, included at least two tasks for each scheme. The tasks and their corresponding schemes were the following: colored tokens (colored tokens); correlations (correlations); chemicals, permutations (combinations); shadows, balance (proportionality); rods, balance (separation of variables); communicating vessels, hydraulic press (mechanical equilibrium). She used the Kendall (1963) $S$ statistic to examine individual behavior for trends in the frequency of success and failure based on a hierarchy of difficulty derived from group performance. She found that for the individual tasks, 35% of the students had response patterns that were consistent with the ordering of the tasks derived from the group mean scores. When the criterion of passing any task within a scheme was used, 60% of the students had response patterns that were consistent with the group ordering.

The performance of each student on the tasks was dichotomously scored "1" to indicate an early or late formal response and "0" to indicate a late concrete or lower level response. From each task response pattern, a scheme response pattern was generated with a score of 1 for a scheme being given if there was a score of 1 for any task measuring that scheme and a score of 0 for a scheme being given if there was a nonformal operational response to each task measuring that scheme. Each student's task response pattern had ten 1's and 0's and each student's scheme response pattern had six 1's and 0's.

Methods from ordering theory were then applied to the data (Bart, 1978; Krus & Bart, 1974). This is a measurement model which identifies necessary, but not sufficient, conditions between test items or tasks. It is classified as a deterministic model which is useful for identifying both linear and branched hierarchical sequences among test items or tasks. It will provide information concerning the following types of relationships:

1. Empirical Prerequisite—A task $i$ is determined to be empirically prerequisite to another task $j$ if a score of 0 for task $i$ does not co-occur with a score of 1 for task $j$.
2. Empirical Equivalence—Two tasks, $i$ and $j$, are considered to be empirically equivalent if the scores on task $i$ are identical to scores on task $j$ for all response patterns.
3. Empirical Independence—A task $i$ is empirically independent of another task $j$ if the score for task $i$ is unrelated to the score of task $j$ (Bart & Airasian, 1974).

The analyses performed were the following: (1) Each set of tasks that measured one scheme (as specified above) were analyzed as to their hierarchical structure. (2) All 10 of the tasks were then analyzed to determine the overall hierarchical structure of the tasks. (3) All 6 of the formal operational schemes were also analyzed to determine their hierarchical structure. The frequency of each response pattern was determined and the hierarchies were identified using specified tolerance levels.

The use of a tolerance level allowed for a compensation for error by removing task response patterns that occurred infrequently. For example, at a tolerance level of 1, all task response patterns which occurred only once were removed and ordering diagrams were constructed for the remaining response patterns according to the rule that task $i$ is empirically prerequisite to task $j$ only if no subject scores 0 for task $i$ and 1 for task $j$ (Bart & Airasian, 1974).

An order analysis, the ordering theoretic counterpart of multidimensional scaling (Krus & Bart, 1974), was then performed on the tasks, and an order-loading matrix was determined for the tasks. A second order analysis was performed on the schemes with an order-loading matrix being determined for the schemes. Order analysis was used in this study to subdivide the sample into subsamples. In other words, linear hierarchies (orders) among the tasks (or schemes) determined by subsamples of subjects, were established as discrete variables analogous

Table 1
Response Patterns for Six Formal Operational Schemes
Determining the First Order Factor

| Scheme | Response Pattern | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Colored Tokens | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Correlations | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Combinations | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| Separation of Variables | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| Proportionality | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Mechanical Equilibrium | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Frequency | 11 | 9 | 8 | 5 | 8 | 5 | 8 |

to continuous factors or latent traits for the tasks (or schemes) and as suggested orders of acquisition of the tasks (or schemes) for subjects within the subsamples.

## Results

Order analysis requires the construction of a subject dominance matrix which provides frequencies of 10 and 01 response patterns for specific pairs of subjects across all tasks (or schemes). This matrix is then used to establish a hierarchy among subjects. The subject dominance matrices for the order analyses of the tasks and the schemes were a 52 × 52 matrix and a 25 × 25 matrix, respectively, and thus were too large to be displayed in this study. However, to illustrate elements of the order analysis procedure, the order analysis of the schemes will be discussed.

From the set of 25 distinct scheme response patterns that were obtained from the 80 students, 7 were selected which determined the longest branch of the subject hierarchy used in the order analysis of the schemes. Those 7 scheme response patterns are presented in Table 1 and also determined the linear hierarchy among the schemes that was established by the largest subsample of students. From the scheme response patterns in Table 1, the scheme dominance matrix was computed and presented in Table 2; that matrix provides the frequencies of 10 response patterns for pairs of schemes across scheme response patterns. The row totals in that matrix determine the order loadings relating tasks to the first-order factor.

Table 2
Dominance Matrix for Six Formal Operational Schemes

| Scheme | | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|---|
| Colored Tokens | 1 | 0 | 5 | 13 | 18 | 27 | 36 | 99 |
| Correlations | 2 | 0 | 0 | 0 | 5 | 13 | 22 | 40 |
| Combinations | 3 | 0 | 8 | 0 | 13 | 21 | 30 | 72 |
| Separation of Variables | 4 | 0 | 0 | 0 | 0 | 8 | 17 | 25 |
| Proportionality | 5 | 0 | 0 | 0 | 0 | 0 | 9 | 9 |
| Mechanical Equilibrium | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 3 provides the entire order analysis results for the schemes and indicates that 67.5% of the sample supported one general order factor and that four group order factors accounted for only 23.8% of the sample. Figure 1 provides three hierarchies for the schemes. The first hierarchy was obtained for a tolerance level of 0 and accounted for 100% of the sample of subjects. The second hierarchy related to a tolerance level of 1 and accounted for 82.5% of the sample. The third hierarchy was obtained for a tolerance level of 2 to 4 and accounted for 67.5% of the sample. For a tolerance level of 0, no prerequisite relations between schemes were evidenced in the scheme hierarchy; however, a perfect linear hierarchy appeared for the tolerance level of 2 to 4, with colored tokens being a prerequisite scheme to all the other schemes.

Additional results suggest that the tasks within some schemes were empirically equivalent. With respect to the proportionality tasks—shadows and balance—77.5% of those task responses were 00 and 11; and the phi correlation between the tasks was .38, a value significant at the .05 level. Both of these results suggest that the two tasks were empirically

equivalent. Also, 90% of the sample was accounted for by one order factor, thus suggesting that there may be only one main order.

Regarding the mechanical equilibrium tasks—vessels and hydraulic press—76% of those task responses were 00 and 11, but the phi correlation between the tasks was a nonsignificant .16. These two results suggest that it is uncertain as to whether those two tasks were empirically equivalent. However, 97.5% of the sample was accounted for by the first-order factor.

The rods and pendulum tasks (separation of variables) appeared to be empirically equivalent, with equivalence being supported by 75% of the sample and with the phi correlation between the tasks being .48, a value significant at a .05 level. Also, 90% of the sample was accounted for by the first-order factor. The combinations tasks—chemicals and permutations—engendered data that was less indicative of an empirical equivalence between tasks than the other three pairs of tasks. Only 55% of the response patterns were 00 and 11, and the phi correlation for these two tasks was a nonsignificant .12 value; but 84% of the sample was accounted for by one or-

Table 3
Order-Loading Matrix for Six Formal Operational Reasoning Schemes
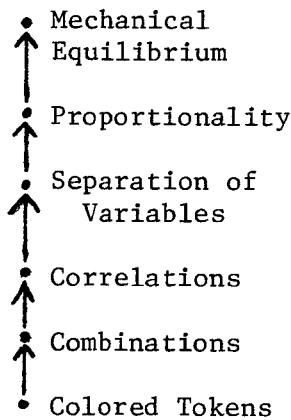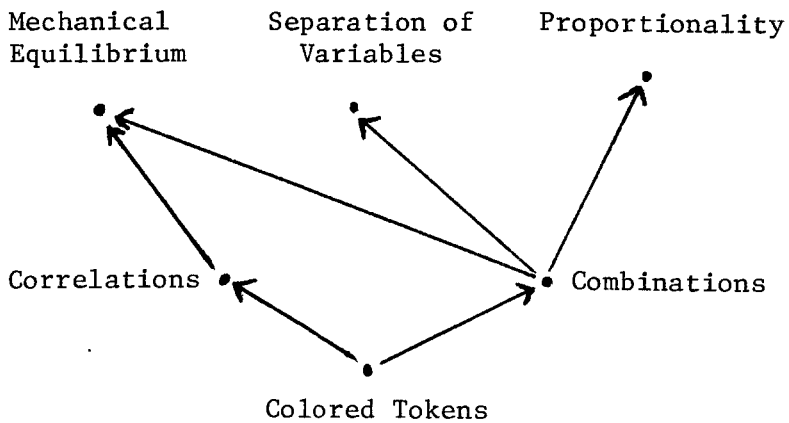
| Scheme | Factor | | | | Row Total |
|---|---|---|---|---|---|
| | I | II | III | IV–IX* | |
| Colored Tokens | 99(.22) | 6(.01) | 10(.02) | 32(.07) | 145 |
| Correlations | 40(.09) | 6(.01) | 4(.01) | 23(.05) | 73 |
| Combinations | 72(.16) | 10(.02) | 15(.03) | 20(.05) | 117 |
| Separation of Variables | 25(.06) | 2(.00) | 0(.00) | 29(.07) | 56 |
| Proportionality | 9(.02) | 0(.00) | 10(.02) | 11(.02) | 30 |
| Mechanical Equilibrium | 0(.00) | 15(.03) | 0(.00) | 7(.02) | 22 |
| Column Total | 245(.55) | 39(.09) | 39(.09) | 122(.28) | 445 |
| Percent of Sample | 67.5% | 7.5% | 6.3% | 18.7% | 100% |

*The Column Totals for Factors IV–IX would average 20.3 and the average Percent of Sample for those remaining six Factors would be 3.1% with none of those Factors accounting for more than 5% of the Sample.

**Figure 1**

Task hierarchies for 6 formal operational reasoning
schemes for tolerance levels of 0, 1, and 2 to 4

- Mechanical
  Equilibrium

- Proportionality

- Separation of
    Variables

- Correlations

- Combinations

- Colored Tokens

Mechanical        Separation of        Proportionality
Equilibrium          Variables

Correlations                              Combinations

Colored Tokens

- Mechanical
  Equilibrium

- Proportionality

- Separation of
    Variables

- Correlations

- Combinations

- Colored Tokens

der factor, suggesting that there was only one order factor underlying the two tasks.

When all tasks were analyzed, the result was the first hierarchy in Figure 2, which held among 6 of the 10 tasks for 100% of the sample and for a tolerance level of 0. The second task hierarchy is displayed in Figure 2, with all response patterns with a frequency less than or equal to 1 being deleted (i.e., tolerance level of 1). This second hierarchy incorporated 52.5% of the sample and was dense, with all 10 tasks being involved in an array of 15 prerequisite relationships. It is interesting to note the absence of empirical equivalences between tasks that assess the same schemes. At best, there were prerequisite relationships between such tasks, e.g., between the balance task and the shadows task. One general order factor accounting for 37.5% of the sample and three group order factors accounting for 23.8% of the sample are indicated in Table 4.

The order analyses tended to reveal sizeable general order factors. In the case of just a few items, tasks, or schemes, it would be difficult to show evidence for more than a single order factor because of the overwhelming influence of the proportion correct. But as the number of items, tasks, or schemes to be analyzed is increased, the secondary order factors will tend to account for larger percentages of the samples. A

**Figure 2**

Task hierarchies for 10 Inhelder-Piaget tasks assessing
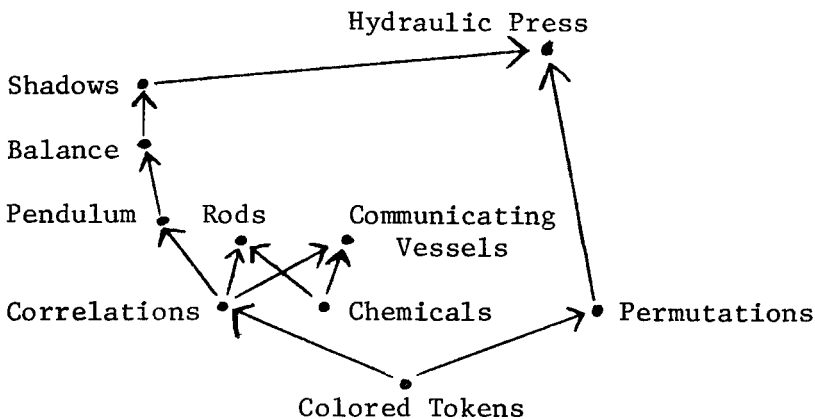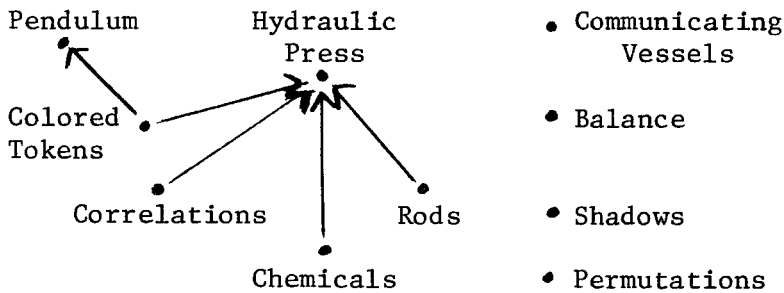formal operational reasoning for tolerance levels of 0 and 1

Table 4
Order-Loading Matrix for Ten Inhelder-Piaget Tasks Assessing
Formal Operational Reasoning

| Task | Factor | | | | Row Total |
| | I | II | III | IV–XVII* | |
|---|---|---|---|---|---|
| Colored Tokens | 112(.08) | 54(.04) | 23(.02) | 164(.12) | 353 |
| Correlations | 67(.05) | 17(.00) | 30(.02) | 99(.07) | 213 |
| Chemicals | 51(.04) | 63(.05) | 38(.03) | 63(.05) | 215 |
| Permutations | 30(.02) | 30(.02) | 0(.00) | 102(.07) | 162 |
| Pendulum | 24(.02) | 3(.00) | 11(.01) | 67(.05) | 105 |
| Rods | 20(.01) | 23(.02) | 11(.01) | 83(.06) | 137 |
| Shadows | 8(.01) | 0(.00) | 4(.00) | 38(.03) | 50 |
| Balance | 14(.01) | 0(.00) | 7(.01) | 41(.03) | 62 |
| Communicating Vessels | 2(.00) | 7(.00) | 47(.03) | 18(.01) | 74 |
| Hydraulic Press | 0(.00) | 0(.00) | 0(.00) | 6(.00) | 6 |
| Column Total | 328(.24) | 197(.14) | 171(.12) | 681(.50) | 1387 |
| Percent of Sample | 37.5% | 12.5% | 11.3% | 37.7% | 100% |

*The Column Totals for Factors IV–XVII would average 42.9 and the
average Percent of Sample for those remaining 14 Factors would be
3.8% with none of those Factors accounting for more than 10% of the
Sample.

case in point is that the second order factor for the 10 tasks accounted for 12.5% of the sample, while the second order factor for the 6 schemes accounted for only 7.5% of the sample. More psychometric research is needed to determine exactly the functional relationships among the number of tasks to be analyzed, the difficulties of the tasks, and the order-loading matrix.

**Discussion**

Some of the results tend to confirm Martorano's results that the tasks within schemes are similar in difficulty and tend to be empirically equivalent. This finding is important because the ordering theoretic analysis does not depend on group differences as did the Kendall $S$ statistic.

The use of a tolerance level of 1 for the analysis of all 10 tasks resulted in a loss of 47.5% of the sample, suggesting that the response patterns were quite varied. This confirms the results of other research that used ordering theoretic analysis with formal operational tasks (Bart & Airasian, 1974; Bart & Frey, 1978). This suggests that the developmental patterns for acquisition and maintenance of formal reasoning skills varies among groups of subjects.

However, one general factor did account for 37.5% of the sample, suggesting that some common structure may underlie performance on the tasks, thus supporting Piaget's notion of the integrative structure of the period. Variation in performance may be due to task, instructional, or performance variables that were discussed by Martorano. The analysis of scheme data more strongly supported Piaget's notion of an integrative structure for formal operations; with a tolerance level of 2 to 4, the schemes formed a linear hierarchy and, from the order analysis, the first

order factor accounted for 67.5% of the sample, both of which suggest that formal reasoning schemes are highly structurally interrelated. Whether data are analyzed by task or by scheme certainly affects the results. Further research with formal operations tasks and schemes and with ordering theoretic techniques will provide insights into issues regarding the integrative structure of cognitive skills and the hierarchical sequence among cognitive tasks.

## References

Bart, W. An empirical inquiry into the relationship between test factor structure and test hierarchical structure. *Applied Psychological Measurement*, 1978, *3*, 333–337.

Bart, W., & Airasian, P. The determination of the ordering among seven Piagetian tasks by an ordering theoretic method. *Journal of Educational Psychology*, 1974, *66*, 277–284.

Bart, W., & Frey, S. *Hierarchies among formal reasoning tasks: Educational ramifications.* Paper presented at the meeting of the Jean Piaget Society, Philadelphia, 1978.

Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence.* New York: Basic Books, 1958.

Jackson, S. The growth of logical thinking in normal and sub-normal children. *British Journal of Educational Psychology*, 1965, *35*, 255–258.

Kendall, M. *Rank correlation methods.* London: Griffin, 1963.

Krus, D., & Bart, W. An ordering theoretic method of multidimensional scaling of items. *Educational and Psychological Measurement*, 1974, *34*, 525–535.

Lovell, K. A follow-up study of Inhelder and Piaget's "The growth of logical thinking." *British Journal of Psychology*, 1961, *52*, 143–153.

Martorano, S. A developmental analysis of performance on Piaget's formal operations tasks. *Developmental Psychology*, 1977, *13*, 666–672.

## Acknowledgments

## Author's Address

Send requests for reprints or further information to William M. Bart, SPPFE, 330 Burton Hall, University of Minnesota, 178 Pillsbury Drive S. E., Minneapolis, MN 55455.