DOCUMENT RESUME

ED 354 147 SE 052 973

AUTHOR	Meyer, Linda A.; And Others
TITLE	The Development of Science Knowledge in Third and Fourth Grade.
INSTITUTION	Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CRS-TR-550
PUB DATE	Mar 92
CONTRACT	NSF-MDR-85-50320; OERI-G0098-C1001-90
NOTE	45p.; For a similar report for kindergarten through second grade, see SE 052 972.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Classroom Techniques; Educational Research; *Family Influence; Grade 3; Grade 4; *Independent Reading; Instructional Materials; Intermediate Grades; Literature Reviews; Primary Education; *Science Education; *Science Instruction
IDENTIFIERS	Process Skills

ABSTRACT

Trends for third- and fourth-grade students were identified from data collected in a longitudinal study of how children learn science concepts. Approximately 325 children from three school districts participated in this study from their first week in kindergarten through the end of sixth grade. A heuristic model of science learning was developed to represent entering ability, home background, home support, lassroom processes, instructional material characteristics, and end-of-year performance. Data were then collected in each of these areas. LISREL analyses were performed to produce a structural model for science learning at each grade level. Results showed that at the third-grade level, only the amount of school activities brought home played an active role in the model. At the fourth-grade level, no measure of home support affected student performance. Classroom instructional variables were endogenous in the third- and fourth-grade models. Independent work made a significant positive difference in the end-of-year performance at the third-grade level but a significant negative difference at the fourth-grade level. In addition, at both levels, general instructional variables affected end-of-year verbal performance but neither time nor other instructional variables (save individual criticism and terminal feedback) made a difference in science learning. Results are discussed generally in terms of the shift away from the significant difference in science performance made by home support after second grade and the impact instruction had on verbal ability but not on science performance. (Author)

CENTER FOR THE STUDY OF READING

Technical Report No. 550

THE DEVELOPMENT OF SCIENCE KNOWLEDGE IN THIRD AND FOURTH GRADE

Linda A. Meyer
James L. Wardrop
C. Nicholas Hastings
University of Illinois at Urbana-Champaign

March 1992

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center, and in part by the National Science Foundation under Grant No. NSF MDR 85-50320 and by grants from the Silver Burdett Ginn and Houghton Mifflin publishing companies, and most recently, by anonymous donors. The publication does not necessarily reflect the views of the agencies and donors supporting the research.

Abstract

Trends for third-and fourth-grade students were identified from data collected in a longitudinal study of how children learn science concepts. Approximately 325 children from three school districts participated in this study from their first week in kindergarten through the end of sixth grade. A heuristic model of science learning was developed to represent entering ability, home background, home support, classroom processes, instructional material characteristics, and end-of-year performance. Data were then collected in each of these areas. LISREL analyses were performed to produce a structural model for science learning at each grade level. Results showed that at the third-grade level, only the amount of school activities brought home played an active role in the model. At the fourth-grade level, no measure of home support affected student performance. Classroom instructional variables were endogenous in the third- and fourth-grade models. Independent work made a significant *positive* difference in end-of-year performance at the third-grade level but a significant *negative* difference at the fourth-grade level. In addition, at both levels, general instructional variables affected end-of-year verbal performance but neither time nor other instructional variables (save individual criticism and terminal feedback) made a difference in science learning. Results are discussed generally in terms of the shift away from the significant difference in science performance made by home support after second grade and the impact instruction had on verbal ability but not on science performance.

THE DEVELOPMENT OF SCIENCE KNOWLEDGE IN THIRD AND FOURTH GRADE

A generation ago, when the Soviet Union launched Sputnik, Americans overnight became concerned about our nation's overall status in science. The response to this concern was for the federal government to invest millions of dollars in the development of innovative programs to teach science. Science educators called for public school teachers to use more hands-on activities to teach science, and some teachers attempted to do so The results, however, were not as hoped for. In fact, Mechling and Liver (1983) described this effort as a "lost sojourn into the abstract and difficult" (p. 43). That American science education is still falling short is evident in surveys that show American students performing poorly on international tests of general science ability. Based on these surveys, Tifft has concluded that "American science education is a shambles" (p. 68).

We argue that a repetition of the post-Sputnik cycle is once again threatening American education. Science educators continue to advocate hands-on science activities. In fact, an entire issue of *Phi Delta Kappan* (May 1990) was devoted to this topic. Our concern is that history may now repeat itself if we have an exclusive push toward hands-on science instruction without first learning how teachers are currently teaching science and how children at different grade levels fare with that instruction.

Why do American children perform so poorly in science? What is the state of American science education? What kind of instruction is being used in typical science classes, and how does that instruction mediate students' science learning over time? It was to address these questions that we undertook our longitudinal study (see Meyer, Wardrop, & Hastings, 1992). This report focuses upon our investigation of science instruction and science learning in third- and fourth-grade classrooms.

Our earlier report (Meyer et al., 1992) presents research related to science learning as well as general aspects of schooling and their impact upon children's learning. This report will focus on our reasons for studying science instruction and science learning at the third- and fourth-grade levels, our hypotheses, and the results of the study.

Research Questions

The purpose of our research is to study science instruction in a naturalistic setting and to see how that instruction and other aspects of third and fourth graders' lives influence what they learn about science. We believe that it is critically important to be able to describe current science instruction and other aspects of children's lives and to correlate these variables so that we can approach an understanding of how American children are developing science knowledge before we suggest reforms.

It is our hypothesis that schooling will explain variance in children's scientific knowledge in the early elementary grades only as a function of the amount of time teachers spend in science instruction while interacting with the children about their background knowledge, giving demonstrations with manipulative materials, asking application questions, having the students formulate hypotheses, and asking other questions that center upon text that are both explicit and implicit. We predict that the most successful elementary grade science instruction will incorporate hands-on *and* text-based instruction. In the absence of such instruction, student achievement in science will be predicted by the children's entering ability and home background variables.

Heuristic Model of Science Knowledge Development

The simplest way to think about the development of science knowledge is to view it as a function of the aptitudes and abilities children possess as they enter school. That is, what and how much children learn about science in the early grades directly reflects their aptitude or verbal ability. A somewhat more complex view sees learning as a reflection both of children's aptitudes and of home-related factors, such as socioeconomic status and the amount and kind of stimulation provided by parents. Science knowledge development can also be seen as a function of the amount and kind of instructional activities children receive in school. In addition, students may be viewed as contributing to their own knowledge development through the reading they do, the television programs they watch, and the activities they choose.

These separate formulations, when combined, more accurately model the nature of science knowledge development. That is, science knowledge development is a function not only of aptitude and of immediate support and instruction, both from home and school, but of student-initiated activities, such as independent reading.

To address the question of how children develop science knowledge over time, we developed a heuristic model to guide data collection and analyses at each grade level (see Figure 1).

[Insert Figure 1 about here.]

The model is composed of eight constructs: home background characteristics, student ability at the time they began school, the characteristics of instructional materials used to teach science, teachers' management style, teachers' instructional style, home support for science knowledge development, student ability at the end of each year, and independent reading.

The following discussion explains how we conceptualized each construct.

1. **Home background.** This construct represents the variables of parental occupation and education, the number of adults in the home, the number of older and younger siblings, and the number of hours each parent works outside the home each week.

2. **Ability, 0.** This construct represents children's verbal abilities upon entering school that are most likely to affect their science knowledge at the end of kindergarten. Ability 0 on the model represents the children's abilities at the first testing in the fall.

3. **Materials.** This construct represents the characteristics of instructional materials that may contribute to children's science de· -'opment. Specifically, it represents the variables of text book content, use, and "considerateness"--that is, the number of problems in the way science textbooks present material.

4. **Management style.** Because we believed that teachers do not necessarily manage their classrooms in the same ways they instruct, we separated teaching initially into two constructs, management style and instructional style. Management style captures teachers' strategies for molding students' general behavior. It is composed of five classroom characteristics: (a) the amount of time teachers allocated to science instruction; (b) their general praise statements to individual students; (c) their general praise statements to groups of students, such as, "Everyone is working very nicely"; (d) their critical statements directed to individual students, such as "Johnny, sit down and start to work now"; and (e) their critical statements directed to groups of students.

5. **Instructional style.** This is the second construct representing teaching. The variables of the instructional style construct are extensions of characteristics reported in research on general teaching effectiveness in other areas of elementary education such as reading and math that have demonstrated the effects that instruction can produce in these areas. Six additional classroom process variables compose this construct. Half of these variables are measures of the kinds of interactions teachers initiated with individual students or entire classes. The remaining half of these variables capture teachers' responses to students who have made errors or who can not come up with an answer. Instructional style was also characterized by the kinds of feedback--sustained, terminating, or confirming--that teachers give.

6. **Home support for science knowledge.** This construct contains three clusters of variables: (a) a child's involvement in science processes with parents, (b) the frequency with which parents provide activities for their children, and (c) the prevalence of science-related books and magazines in the home to which the child has direct access.

7. **Independent reading.** This construct represents reading initiated by the child. We anticipated that activities such as independent reading might influence children's science knowledge in later grades, but probably not in the kindergarten through second-grade years.

8. **Ability, 1.** This construct represents students' science knowledge in the spring of each school year.

Method

The Setting

All of the students, parents, and teachers at designated grade levels in three school districts have taken part in this research. In each district, we have studied two cohorts (age groups) of students. Two of the districts had only one elementary school. The third district had 10 elementary schools but only one of these participated in the study. These three districts were selected because of the natural variations they provide in educational philosophies and approaches and instructional materials. We will show how these differences in approaches were reflected in differences in time and other instructional characteristics of science teaching.

District A. District A is in a fairly self-contained small town in the Midwest. Students in the school studied have mixed abilities upon entering kindergarten. There were approximately 90 children per cohort in three classes at most grade levels. This district is known for its high student performance in reading and its average student performance in science in the early elementary grades. The district's educational philosophy includes whole-class instruction in all subjects beginning with an academic kindergarten. Children are almost never grouped for instruction in District A, as district policy dictates that all students at a grade level will cover the same content each year. Regular kindergarten and first-grade classroom teachers in this district maintain primary responsibility for children with special needs in their classes, and often gear many of their instructional interactions and feedback to the lower performing students. Merrill's *Accent on Science* (Sund, Adams, & Hackett, 1982) is the text adopted districtwide.

District B. District B is in a small town near a larger town in the Midwest to which many of its citizens commute to work. The school from this district accounts for about 150 children in Cohort 1. The district has a tradition of average student achievement in reading and better-than-average performance in science in the lower elementary grades. Teachers here begin grouping for some instruction in kindergarten, and this practice continues through all of the elementary grades. Classes are divided into as many as five or six groups for reading instruction in first and second grade. Thus, students in these classes, on average, spend considerable time each day in independent work. Special teachers begin to play important roles with lower performing children in first grade. There is also some tracking of students into self-contained classes even at the first-grade level.

The year Cohort 1 was in second grade saw dramatic changes in District B's science instruction. Prior to that year, the district had no textbook for science instruction in the lower elementary grades. In 1985-86 the district piloted the Holt science program (Abruscato, Fossaceca, Hassard, & Peck, 1980), and a regular classroom teacher was appointed at each grade level to coordinate science instruction. Each coordinating science teacher had the responsibility of collecting materials teachers needed to implement the activities and experiments in the Holt program. This teacher leader also distributed materials to all other teachers at the grade level. The teachers also used their science program's tests on a regular basis to evaluate student performance in science.

District C. The school participating from District C, located in a suburb of a major Midwestern city, has many characteristics of an urban school. It is recognized as the highest performing elementary feeder school to its middle school. It has a heterogeneous population, with students from mixed socioeconomic and ethnic backgrounds. White, Black, and Hispanic children made up this portion of the sample. There were approximately 85 children per cohort in three "classes" that were grouped as combined grade-level homerooms during first and second grade and again during fourth and fifth grade. Kindergartens were self-contained. Special teachers played a very important role in this school in District C. Bilingual instruction began in kindergarten for children whose parents chose it. These children were pulled out of their regular classes for about 30 minutes each morning and again each afternoon for instruction in English and Spanish. Teachers worked with "low stanine" children as a separate group from first grade on.

Science was taught in homerooms. For first and second graders, these homerooms spanned two grade levels. Teachers' committees used the Silver Burdett (Mallinson, Mallinson, Smallwood, & Valentino, 1985) science textbooks at each grade level to develop units for instruction in homerooms. Science lessons for first and second graders were taken from the first- and second-grade books. During actual instruction in kindergarten through second grade, teachers most often concentrated on verbal presentations of science concepts with a strong emphasis on students' background knowledge. Teachers also focused on activities and experiences during these lessons. They seldom referred to science textbooks in either first or second grade, and therefore the textbooks were almost never seen by us in use in these classrooms.

As these short descriptions reveal, there was substantial natural variation within each of these districts. Each of the districts is fairly typical of numerous school districts in the United States, and together they yield a composite of characteristics that typify American elementary schools today.

Measures

To track the development of children's science concept acquisition, we have administered a number of psychometric measures. While we have given some standardized measures of verbal performance and basic science understanding, we have found that such published tests do not entirely satisfy the needs of this study. Therefore, we developed a number of customized measures of verbal performance and basic science knowledge with algorithms specifically for use within this context. The following description of the standardized measures of verbal performance and basic science knowledge used in the study will be brief because detailed accounts of each can be found in the manuals and reports provided by their publishers. The description of the custom-developed measures will also be brief, because they have been described in some detail elsewhere (Hastings, Meyer, & Linn, 1987; Hastings, Meyer, Linn, & Wardrop, in press). Table 1 presents means and standard deviations for the sample on each measure.

[Insert Table 1 about here.]

Standardized Measures of Verbal Performance

While the primary thrust of this research was to account for variance in children's acquisition and application of science concepts as opposed to their learning of rote scientific facts and vocabulary, we believed that a certain level of verbal competence was necessary for children to understand basic science concepts. Therefore, we included several measures of verbal-reading performance in our model. Several of these were standardized tests that have been nationally normed.

Wide Range Achievement Test. We administered the reading subtest, Level I of the Wide Range Achievement Test (WRAT), at least once a year during the course of this study. The children participating in this study received the 1984 version (Jastak & Wilkinson, 1984) beginning in third and fourth grade. Items on the WRAT consist of a series of increasingly difficult words that children read aloud to the examiner. The measure is individually administered and has a stopping rule whereby 12 consecutive errors terminate administration. Fall WRAT scores were used as one variable for the latent trait Beginning Verbal Performance at each grade level.

Woodcock Reading Mastery Tests. The reading comprehension passage subtest of the Woodcock Reading Mastery Tests (Woodcock, 1973) was administered at least once a year to children in the longitudinal study. The basic format of this subtest consists of increasingly difficult cloze passages children read. A stopping rule based on five consecutive errors is used to terminate administration. Like the WRAT, the Woodcock was used as an observable measure of the latent trait Beginning Verbal Performance at each grade level.

Customized Measures of Verbal Performance

Although we used standardized tests of verbal performance, we felt that such tests did not precisely measure the latent traits we wished to study. We, therefore, used a customized measure developed by other researchers.

Interactive Reading Assessment System (IRAS). The Interactive Reading Assessment System (IRAS) (Calfee & Calfee, 1982) asks students to read up to 14 word lists of 8 words each until the stopping rule based upon accuracy is applied, or until the last list is read. Rate, accuracy, and self-corrects were recorded. Students then read up to seven passages of increasing length and difficulty until they made more than 10 decoding errors and missed half or more of the comprehension questions. Rate, accuracy, and self-corrections were recorded for this section also. Correctness of responses to questions based on the passages with or without a prompt was recorded as well. This measure was used for LISREL modeling as a manifest trait for beginning of third-grade verbal performance.

Customized Measures of Basic Science Knowledge

We developed instruments to test students on content domains from their science textbooks that were either at or above grade level. We tried to use vocabulary common to all three school districts, and we included out-of-level items in each battery so that we could observe children developing knowledge on a few select science topics.

Motion Test. The Motion Test (Meyer, Hastings, & Linn, 1986) is administered orally and is composed of items exclusively out-of-level for third graders. It includes items with balls and mirrors and other objects. The children are asked to identify the paths these objects would take to or from a flat surface. We used the Motion Test initially as an end-of-second-grade manifest variable for basic science understanding. We used it again at both the third- and fourth-grade levels.

Living Things Test. The Living Things Test (Meyer, Hastings, & Linn, 1987a) has in-level and out-of-level items, common vocabulary, and items represented by line drawings whenever possible. This test includes items on plants that are used as out-of-level items at the second-grade level. It also includes items on animals since "animals" represents the only other content domain common to the three school districts participating in the study.

Science Passage Test. The Science Passage Test (Meyer, Hastings, & Linn 1987b) was developed by selecting passages with content foreign to all curricula used by the schools participating in this study. This instrument is composed of a passage on muscles and rubber bands and of a passage on the planets taken from the Distar Reading III Program (Engelmann & Hanner, 1972). This test is administered one-on-one by having the children read orally. We ask the children background-knowledge questions and questions related to the content of each passage immediately before having them read it. We ask comprehension questions after the children have read the passages. In addition, we time each child's reading and code oral reading errors. Therefore rate, accuracy, and comprehension scores come from this instrument.

Open Court Test. The Open Court Test (Open Court, 1987) is a measure intended to test children's abilities to identify errors in passages that spoil the meaning of those passages. The test was derived from selections from the Open Court series Catching On (Anderson, Bereiter, & Smart, 1977) with permission from the publisher. Children read these passages silently and then circle a word or phrase that spoils the meaning. Selections range in length from a few short sentences to text more than half-a-page long.

Measures of Verbal Performance and Basic Science Understanding

Of the various measures we gave, this test appears to be suspended somewhere between verbal and/or cognitive performance and measures of basic science understanding.

Sequential Test of Educational Progress Science Subtest. The science subtest of the Sequential Test of Educational Progress (STEP) (Educational Testing Service, 1979) was given to our Cohort 1 children in the spring of their third- and fourth-grade years. The subtest tends to load more with "reading measures" than it does with certain "science concept measures," specifically the Motion Test. This was true for both a promax-rotated forced two-factor analysis (Hastings et al., 1987) and a confirmatory two factor analysis using LISREL (Hastings et al., in press). This may be due to the fact that students must read items on this instrument silently in order to respond.

[Insert Table 1 about here.]

Procedures and Results

Instructional Style

Classroom observations for third and fourth grade were conducted in the manner described in Meyer et al. (1992). Each classroom teacher was observed for six full days. Special teachers were observed four full days each. Observers made scripts of each science lesson. They also noted when each lesson started and ended. Audio tapes were also made for back-up use in case they were needed for coding observations.

All instructional interactions were coded. Therefore, the observational categories were expanded as teachers introduced new types of activities and interactions. Thus, the number of categories grew and expanded as need be. This flexible system allowed for complete and accurate coding of all science instruction. Interrater reliability was always at least .85 in practice and when observers switched classes. In the tables that follow, all frequencies are reported at the student level. In other words, we report the average number of minutes each child had and the average number of teacher-directed interactions and feedback each student received during one of the observations.

Third grade. The results of observations of third-grade classrooms for time spent in science appear in Table 2. There were substantial increases in time spent in science instruction at the third-grade level. Districts A and B averaged over 18 minutes of instruction per observation, while District C students spent close to 15 minutes per observation. Therefore, students in Districts A and B received over 90 minutes of science instruction each week, and students in District C averaged over 70 minutes a week.

[Insert Table 2 about here.]

Table 3 shows results for the five major types of interactions teachers had with students as well as their frequencies for different types of feedback and time children spent working independently. District C teachers averaged a larger number of text-tied, general review, and oral reading interactions in their classes. District B teachers provided the most sustained feedback and the least terminating feedback to their students. This creates a pattern of instruction focused on staying with children who were having difficulty coming up with correct answers. District B teachers had their students work independently decidedly more than the teachers in the other districts. They also were observed covering the largest number of different science content domains.

[Insert Table 3 about here.]

Fourth grade. Time spent in science instruction in fourth grade is reported by district in Table 4. Once again, there were substantial differences from the previous grade level for times spent in science instruction. The greatest gain was found in District C, while the District A teachers actually spent less time in science in fourth grade than they had spent in third grade. The greatest variance between teachers and the highest amount of time were both found in District B.

[Insert Table 4 about here.]

Table 5 presents means and standard deviations for interaction and feedback categories found in fourth-grade classrooms. Only those variables found to be significant in the final LISREL model are reported here. Teachers' uses of review and concept application interactions made significant differences in children's performance in science. Each child averaged more than one general review interaction per observation, and District B students received review questions more frequently than did children in either of the other two districts. Frequencies for concept application interactions were much lower than general review interactions thus suggesting science instruction focused more upon reviewing information that had already been presented than on having students apply new information. District B teachers used these interactions most with their students.

There was a similar pattern for sustained feedback. Teachers in Districts B and C used the most sustained feedback with their students. Terminal feedback was negatively and significantly related to student performance. District C teachers used it the most whereas District A teachers used it very infrequently.

[Insert Table 5 about here.]

Management Style

Third grade. At the third-grade level none of the teachers' management style variables mediated the children's performance in science significantly. Nonetheless, the descriptive frequencies are presented here for comparison purposes. The third graders averaged being on task over 89% (89.91%) of the time when they were working independently. They spent just over 100 minutes in independent work each day observed. Their teachers used both praise and critical statements to individuals and groups sparingly. When individual and class praise statements are summed, the average is still just over 10. These teachers averaged just 5.29 critical statements to individuals and only slightly more (5.65) to classes.

Fourth grade. Unlike the third-grade level where none of the teachers' management style variables mediated children's performance, at the fourth-grade level we found four management variables to mediate performance. The percentage of students on task during independent work time as well as the amount of independent work time students had allocated to them both made differences in their performance in science. District B students were on task most frequently (almost 96% of the time), and they had more time to do independent work than children in the other districts. Both praise statements made to entire classes and criticisms to individuals also mediated student performance. District C teachers praised their classes the most, and District A teachers criticized individuals in their classes more frequently than did teachers in the other two districts. Criticism to individuals was negatively related to student performance in science.

Instructional Materials

Third grade. Textbook use continued into third grade at all three schools. Once again there were substantial differences between the textbooks' characteristics. The Silver Burdett text had 15 content domains and 243 vocabulary words. The Merrill series had 20 content domains and 96 vocabulary words. Holt presented just 6 content domains and specified 137 words to be learned as vocabulary.

Content coverage observed included three forms of matter, machines, and space as well as various lessons on living things, rocks, and health in District A. Many of these same topics were observed in District B. In addition, lessons on space, the earth, heat, the sun and the stars, and dinosaurs were observed in District B. The three District C teachers divided science teaching among themselves so that one teacher taught science units, a second teacher taught health units, and the third presented lessons on living things, magnets, and space.

Fourth grade. Fourth-grade teachers again used the Merrill, Silver Burdett, and Holt textbooks identified first for each district at the second-grade level. The general differences found among these series at the third-grade level were also readily apparent at the fourth-grade level. Silver Burdett had 15 content domains and 288 vocabulary words. Merrill presented 25 content domains and introduced 137 vocabulary words. Holt increased its number of content domains to 11 while introducing 146 vocabulary words.

During the classroom observations, we saw District A fourth-grade teachers present lessons on health, light, outer space, magnets, body systems, and animal behavior. District B teachers taught lessons on weather, machines, the food cycle, light, plants, the earth, and populations. District C teachers were observed teaching lessons on living things, electricity, and mixtures.

Home Support Indices

Descriptive results from the indices are summarized in Table 6. The index of participation in activities, a general measure of how much parents involved children in science-related activities at home, yielded scores that ranged from 0 to 30. Its largest overall mean was at the kindergarten level (Meyer et al., 1992), and it had dropped to just 8.58 in third grade. There was a slight increase in this index for fourth grade (12.21). In contrast, the index of experience with adults, which provides a measure of the frequency of experiences children are engaging in with their parents, rose consistently through the lower elementary grades (from 31.55 to 46.14) and continued to rise in these grades with a mean of 48.15 for third grade and 49.93 for fourth grade.

[Insert Table 6 about here.]

Structural Modeling of Science Knowledge Attainment

In our attempts to account for students' science knowledge attainment at the end of each school year, we employed structural equation modeling using Joreskög's (1978) Linear Structural Relations (LISREL) model. Analyses were performed using the LISREL VI computer package (Joreskög & Sorbom, 1984). A brief overview of the LISREL approach appears in Meyer et al. (1992). Therefore, only a very brief overview follows. For a more detailed discussion of the approach, see Joreskög and Sorbom.

The LISREL model incorporates two components: a structural equations model and a measurement model. The structural equations model, like the more traditional path analysis approach, seeks to estimate structural (path) coefficients for a hypothesized model of the relationships among constructs or variables; the measurement model uses confirmatory factor analysis techniques to accommodate the use of multiple indicators for the latent variables (constructs) of interest and to take into account measurement errors and specific variances (in the factor analytic sense) of the observed variables.

We used the maximum likelihood estimation procedure which provides a chi-square index of goodness-of-model fit, t-statistics for the significance of individual parameter estimates, and modification indices suggesting how much improvement in fit could be obtained by relaxing the various constraints (see below) specified in the current version of a model. We sought a ratio of the chi-square statistic to its degrees of freedom of 2.5 or less; a goodness-of-fit index of .90 or more; and a root mean square residual of .06 or less.

This description of the structural modeling analyses consists of three parts. The first two parts are repeated at both third and fourth grade. An initial model, representing an instantiation of the general heuristic model, is described; then the major modifications made during the analysis are summarized, along with presentation of the "final" model that resulted. The third part is a discussion of the trends observed as we moved from third and fourth grade with respect to the importance of the "classroom" variables vis-á-vis variables characterizing home and pupil influences on end-of-year attainment. The general considerations that guided our work centered upon univariate distributions and multi-collinearity. These concerns are summarized in Meyer et al. (1992).

Third-Grade Analyses

For the third-grade analyses, a more complex model was initially posited to account for end-of-year science performance than initial models for the earlier grades. That is, although we began our analyses with a variation on our general model, there were more variables, somewhat more complicated measurement structures, and more structural connections among variables and constructs to be accounted for. We began analyses with an initial pool of 57 observed variables. Preliminary exploratory analyses, including frequency of occurrences of classroom related behaviors, distributional characteristics of observed variables, correlational analyses, and preliminary factor analyses, led us to drop 11 variables from consideration in the model: (a) hypothesis interactions, (b) manipulation of objects interactions, (c) classification interactions, (d) time the teacher spent manipulating objects, (e) time the students spent manipulating objects, (f) a "catch-all" feedback variable, (g) the number of science domains per textbook, (h) the number of vocabulary words introduced per text domain, (i) the number of propositions per common text domain, (j) the number of questions per domain, and (k) the number of activities per domain. We also used the simple algebraic transformation of obtaining the square root of three variables: background knowledge interactions, criticisms to whole class, and criticisms to individuals. Finally, we summed individual and class praises to obtain one variable, teacher praise. With the remaining 44 variables we were able to achieve a positive-definite correlation matrix with which to begin our initial modeling. This model is represented in Figure 2.

[Insert Figure 2 about here.]

Employing the iterative approach discussed previously we used LISREL to explore alternative specifications of both the measurement and the structural equation components of the model. In this manner we arrived at the final structural model shown in Figure 3. This model retained 36 of the original 44 observed variables as manifestations of eight latent variables and one single indicator variable. It might be noted at this time that in contrast with the kindergarten through second-grade models reported in Meyer et al. (1992), the classroom variables were best considered to be endogenous. That is, children's beginning-of-year performance in science tended to have an effect on teacher behaviors and classroom activities observed during the school year. Therefore, the only variable that remained exogenous to the model was the construct we refer to as home background.

[Insert Figure 3 about here.]

Home Background Variables

As was the case for kindergarten through second grades, what we refer to as home background was defined as a composite of the four parental education and occupational prestige measures. Again, because these measures were obtained for all pupils during their kindergarten year, the structure of the home background variable was similar to previous analyses. Mother's occupational prestige had the lowest representation in this structure (loading = .49) while mother's educational level contributed the most (loading = .85). Father's occupation (loading = .65) and father's education (loading = .78) fell in between. Other home background variables, such as number of hours either parent worked per week, had no significant role in the model.

Home Support

We began analyses expecting that a home support composite variable would be formed by three variables. However, only the amount of school activities that were brought home during the third-grade school year seemed to play an active role in the model. This suggests that home support was linked to school activities. Therefore, the final model contained only a single indicator variable called school activities. This variable had a small but significant effect on both end-of-year performance composites.

Beginning-of-Year Test Performance

For the beginning-of-third-grade cognitive measures, we once again defined two composite variables that represented entry capability in verbal performance and science performance. Verbal Performance was manifested in part by scores on the WRAT (loading = .79) and the four passage-related subscores on the IRAS. The sum of correct responses to passage comprehension questions had a loading of .72, while the relative rate, word errors, and self-corrections loaded negatively (-.88, -.79, and -.49, respectively), a result which was expected because high scores on these three measures would be characteristic of poor readers. Additionally, the Woodcock and the STEP Science subtest each contributed to both beginning-of-year composites. The Woodcock had a loading of .68 on verbal performance, while the STEP Science subtest (administered at the end of second-grade) had a relatively small loading of .38 on this composite. The verbal loading of the STEP is consistent with second-grade end-of-year performance results and may be due to the fact that children needed to read item stems and distractors in order to complete the instrument. It is noteworthy, however, that at the third-grade level, beginning-of-year verbal performance had no direct relationship to final science performance. This suggests that science knowledge and general verbal ability are separate constructs. We are measuring different abilities in each construct, and children high in one are not necessarily high in the other.

The entering science performance composite was defined by seven measures. Four of these--Plants (loading = .73), Three Forms of Matter (loading = .74), Motion (loading = .41), and the STEP Science subtest (loading = .61)--had been administered during the spring of the students' second-grade year and served as proxies to entering third-grade performance. Two CIRCUS subtests, Listening and Think It Through (loadings of .70 and .72, respectively), and the Woodcock (loading = .34) were given during the fall of third grade. The trend from previous years continued, in that beginning science performance was easily the strongest predictor of final science performance (β = .84).

Classroom Variables

The modeling of classroom instructional variables was quite complicated for the third-grade analyses. The final model included three highly overlapping clusters we called General Instructional Activities, Instructional Management, and Independent Study Activities. As noted above, these composites were endogenous in the model. The independent study composite was primarily defined by the frequency of seatwork that students engaged in (loading = .94) and to a lesser extent by the number of domains covered in the materials used (loading = .29). Not surprisingly, interaction and feedback variables had negative loadings on this variable. Children's entering science abilities seemed to have a relatively strong influence on the amount of independent study in science they were assigned during the school year (ß = .25). Independent Study Activities, in turn, had a small but significant relationship to end-of-year science performance (ß = .06, t = 1.71).

The composite General Instruction Activities was defined by eight direct measures, although six of these loaded on at least one of the other two composites. Of the eight, the four that seemed to best define this composite are noted here. Teacher-initiated interactions, which were reviews of material or subject matter that had been previously encountered in the classroom, had a loading of .85 on this variable. The other types of interaction most strongly connected with general instruction were those that were text-tied or based on materials that the students were using during a given science activity. Text-tied interactions loaded .75 on this variable. Terminal feedback (loading = .81) was the dominant feedback given to students during general instruction activities, while sustaining feedback (e.g., giving hints until the student responds correctly) had a somewhat lower loading of .68. General Instruction Activities had a small but significant impact on end-of-year verbal performance (ß = .06, t = 2.53).

We called the third classroom instructional variable Instructional Management. It was best characterized by the amount of time spent in science (loading = .96), the frequency of procedural interactions (loading = .73), and by the number of domains taught from materials being used (loading = .60). This construct had no direct relationship to end-of-year performance but was retained in the model to help account for some of the high degree of overlap between the classroom variables.

End-of-Year Performance

As was the case at the second-grade level, we had two composite performance measures derived from tests given at the end of the third-grade school year. The latent variable we called Verbal Performance was defined by scores on the Open Court (loading = .40) and three subscores of reading errors on individually read science passages, Decoding Errors (loading = -.72), Self Corrections (loading = -.47), and Rate (loading = -.88). In addition, the STEP Science subtest had a loading on the Verbal Performance composite of .20.

The second cluster, which we called Science Performance, was defined by scores on the Motion test (loading = .40), comprehension scores on the science passages (loading = .64), the Living Things test (loading = .62), the Open Court (loading = .41), and the STEP science subtest (loading = .75). It might be noted that both the STEP Science and the Open Court load on Science and Verbal end-of-year composites indicating some degree of overlap between the skills assessed by the instruments.

Structural Model

Structural equation coefficients. The model graphically represented in Figure 3 seems to us to represent a reasonable set of structural relationships among the third-grade variables. As noted above, this model retained 36 of the initial pool of 57 variables (44 variables were used in the somewhat simplified initial LISREL model). Given the complexity of this final model, the fit statistics were quite good: chi-square/df = 2.24, goodness-of-fit index = .821, and root mean square residual = .077.

The home background composite was the only variable that remained exogenous to the third-grade model. It had a strong positive effect (.46) on beginning-of-year science performance, a somewhat weaker effect (.37) on beginning-of-year verbal performance, and a relatively small but significant effect (.11) on school-related activities that were brought home.

Not surprisingly, beginning-of-year science performance once again had the strongest relationship to final science performance (ß = .84). The effect of beginning-of-year verbal performance on final science performance finally dropped out of the model altogether, except for indirect influence through final verbal performance which did have a significant effect (ß = .18). The effect of initial verbal performance on final verbal performance was quite strong (ß = .94). School activities taken home had a small but significant direct effect on final science performance (ß = .07, t = 1.722) and a negative (and not significant) effect on final verbal performance (ß = -.05, *t* = -1.778).

A major distinction of the third-grade model from models obtained in previous years is that classroom activities were endogenous at this level. That is, students' initial science performance had a direct effect on the cluster of variables that made up the latent variable we called Independent Study Activities (ß = .25), which in turn had a small but significant effect on final science performance (ß = .05, *t* = 1.711). The classroom variable General Instruction Activities also had an effect on final verbal performance (ß = .06).

Total effects on final performance in science. Table 7 presents a summary of the total effects found in this final model. As was previously mentioned, the home background composite remained exogenous to the model. This composite had an indirect effect on end-of-year science performance of .48. While beginning-of-year verbal performance had no direct effect on end-of-year science performance, there was a total effect of approximately .17. End-of-year verbal performance had a total effect of .19. Clearly the strongest effect on final science performance was, once again, entry-level science performance (1.036) Though not a direct influence on end-of-year measures, it is noteworthy that beginning-of-year science performance had a total effect of .28 on independent study activities which in turn had an effect of .05 on final science performance.

[Insert Table 7 about here.]

Fourth-Grade Analyses

As we have moved up through the grades, the models that seem to account satisfactorily for science achievement have become increasingly complex: more variables, more complicated measurement structures, more structural connections among variables and constructs. We began this analysis of fourth-grade performance with 52 variables, using as our initial model a close approximation to the final model for the end of third grade (Figure 3). After eliminating some and algebraically combining others to achieve a positive-definite correlation matrix, our initial LISREL analysis involved 44 variables.

Using the same overall strategy as for previous grades, we arrived at the model portrayed in Figure 4, in which 38 variables remain. As before, we first summarize relationships in the measurement models for exogenous and endogenous variables, then we describe the structural model used to account for interrelationships among constructs.

[Insert Figure 4 about here.]

Measurement Models

Home background and home support variables. In contrast to results from previous grades, only stable home-based variables were retained. Among the variables presumed to be stable over time, the education/occupation indicators continued to cluster as indicators of the Home Background composite. At this level, the only other background variables that had any significant and meaningful relationship with the rest of the model were Mother's Hours Worked per Week and Father's Hours Worked per Week. None of the home support variables had a significant relationship to end-of-fourth-grade science achievement. One way of interpreting this result for the Home Support composite is as a logical extension of the trend, observed previously, for the influence of home based activities to have less and less effect on school achievement. The total effect of Home Background on end-of-fourth-grade science achievement was .50.

Classroom observation variables. Only 10 of the classroom-based variables remained in the final model for fourth grade. (See Table 8 for a detailed explanation of which variables were eliminated, and why.) These 10 variables formed two overlapping clusters, plus two stand-alone indicators. The first cluster seems to be characterized by activities and interactions emphasizing instructional tasks and is anchored by the text-based teaching variable, with strong loadings from science instructional time and application interactions as well. The second composite seems to be primarily focused on the allocation of class time to individual seatwork (frequency of seatwork loads .93 on this composite), with smaller positive contributions from application interactions (.59) and text-based teaching (.37) and a relatively strong negative loading (-.70) for teacher praise directed to the whole class. The two stand-alone indicators are individual criticism and terminal feedback. For both of these observed variables, the reliability was (arbitrarily) fixed at .81.

Beginning-of-year performance tests. With the inclusion of measures from both the end of second grade and the beginning of third grade to have a more complete representation of students' skills than would have been possible with only the latter, the "beginning-of-year" tests formed two composites, one that seemed to emphasize reading skills and a second whose interpretation was less clear and has (only tentatively) been labeled Science Performance, Beginning 4th. All our measures of science knowledge load only on this composite, although several of the presumed "reading" tests (specifically, the WRAT, Woodcock, Open Court, and STEP Reading) also have significant loadings on this composite, as does the STEP Listening subtest. The correlation between these two composites is not fully accounted for by those antecedent variables in the model, so that the residuals from the two had a covariance of .38 (representing a correlation of about .46).

Eight measures comprised the Verbal Performance composite, with loadings ranging from .28 for the Open Court test to .93 for the Science Passages Reading Rate. (The three variables that represented error-based measures and therefore would be expected to have negative loadings were recoded so that their expected loadings would be positive.) Two of the Science Passages variables (Reading Rate, with a loading of .93, and Decoding Errors, with a loading of .82) dominated this composite.

The second composite was defined primarily by the Living Things Test, with a loading of .69; the Science Passages Comprehension score, with a loading of .63; and the STEP Science subtest, with a loading of .61. As noted above, three of the standardized reading tests had significant loadings on this factor. Once again, the Motion test has a much smaller loading than do the other science measures on this predominantly science composite.

End-of-year tests. The measures obtained at the end of fourth grade formed two composites that we have called Verbal Performance, End 4th and Science Performance, End 4th. Two of the error-based measures from the Science Passages (Decoding Errors, with a loading of .68; and Self-Corrections, with a loading of .72) had the largest loadings on the verbal performance composite, with Science Passage Reading Rate, the Open Court items, and STEP Science also included. The Science Performance composite was defined primarily by two specially constructed science tests (Living Things and the NSF Test, both with loadings of .76) and the STEP Science subtest (.73), with smaller loadings for other putatively science measures.

Structural Model Components

More clearly than in any previous analysis, it was necessary to treat classroom observation constructs (composites) as endogenous variables, affected by beginning-of-year pupil performance (paths from beginning-of-year verbal or science performance to every classroom-based construct). More will be said about these relationships in sequel.

Home-based variables. The home-background composite had significant positive effects on both beginning-of-year performance composites, and each of the hours worked variables influenced one of these performance composites. Home background had a substantial positive effect (.53) on beginning-of-year science performance, and a smaller but still appreciable effect (.22) on verbal performance. Inexplicably, mother's hours worked per week had a positive effect (.08) on beginning-of-year verbal performance, indicating that children whose mothers worked more hours per week had slightly greater skills in this area. This relationship, although statistically significant, was quite weak. In contrast, the effect of father's hours worked per week on science performance (-.15) was consistent with our intuitive expectations: the more the father worked (and, presumably, the less he was available at home), the lower the child's beginning-of-year science performance.

Entry-level achievement. The beginning-of-fourth-grade verbal performance composite had, as expected, a strong positive effect on end-of-year verbal performance (ß = .95), and it also had significant negative effects on three classroom variables: teacher criticisms to individuals (ß = -.06, indicating a slight tendency for teachers to direct fewer criticisms to the higher-achieving pupils), terminal feedback (ß = -.17, as teachers used terminal feedback more often with lower achievers), and the seatwork activities composite (ß = -.17, indicating that lower achievers received more application interactions and time in individual seatwork activities than did higher achievers. Additionally, given the moderately large negative loading of teacher praise directed to the class as a whole on this composite, such praise tended to be given more in classes where the average entry-level verbal performance was higher). It must be emphasized that these are very weak relationships, and that 97% of the variance in the Seatwork Activities composite was unexplained in this model.

Beginning science performance had a direct positive effect on the Instructional Activities classroom composite (ß = .34), indicating that pupils with higher beginning-fourth-grade science skills engaged in more activities associated with text-based teaching, spent more time on science, received more application and general-review interactions and sustaining feedback, and had more time on task during seatwork. Beginning science performance also had a significant negative effect on teacher criticisms directed to individuals (ß = -.24), suggesting that criticisms were directed more to the lower-achieving pupils. Finally, and predictably, entry-level science performance had a very strong positive effect on end-of-year science performance (ß = .93).

Other influences on end-of-fourth-grade science achievement. The direct effects of beginning-of-year performance on end-of-year achievement have already been described in the preceding section. What remains is to note how classroom process variables influenced final performance. The Instructional Activities composite had no direct effect on science achievement, a result we find most puzzling. It did have a weak negative effect on verbal achievement (ß = -.08).

The two individual indicator variables (individual criticisms and terminal feedback) and the Seatwork Activities composite had weak but significant effects on end-of-year science achievement. Appropriately, individual criticisms had a negative effect on final performance (ß = - .10), but the other two relationships--terminal feedback to science performance (ß = .11) and Seatwork Activities to science performance (ß = .11)--are counterintuitive and, indeed, unreasonable.

Although this model accounts for 94% of the variance in end-of-year science achievement (with beginning-of-year science performance alone accounting for about 87%), we consider it problematic.

Total effects on final performance in science. As Table 8 indicates, home background had a substantial (indirect) effect on end-of-year performance, .50, and the hours per week the father worked had a moderate negative (indirect) effect of -.15. As in previous grades, by far the largest effect on science achievement at the end of fourth grade was beginning-of-year performance, with a total effect of approximately .95. Verbal performance had almost no effect (.01 for beginning verbal performance, .00 for final verbal performance). For the classroom process variables, instructional activities (NSTRATV) had no effect, independent study activities (INDSTATV) had an inexplicable negative effect (-.11), terminal feedback from the teacher had a modest positive effect (.10), and teacher criticisms directed at individual pupils (INDCRIT4) negatively affected end-of-year performance (-.11), as might logically be expected. The pattern of effects presented in Table 8 and described here contributes to our questions about this model. These results are counterintuitive, to say the least.

[Insert Table 8 about here.]

Highlights of the Fourth-Grade Structural Model

Several interesting developments are worth noting in connection with this model. First, the influence of students' entry-level skills on teacher classroom behavior continues, although with implausible effects on end-of-year performance. The path from Beginning Science Performance to Instructional Activities has already been discussed in the previous section and seems noteworthy. Distressingly, this cluster of teacher behaviors did not show a significant relationship to end-of-year science achievement.

Also, home background and the other stable home characteristics are the only exogenous variables in this model. These home-based measures influence end-of-year science achievement only indirectly through their effects on entry-level science performance.

Finally, the measurement structure of our assessments of beginning-of-fourth-grade achievement indicated two factors: one primarily verbal, with a special emphasis on what appeared to be "decoding" skills; the other primarily science, but with considerable overlap with some of the verbal-comprehension measures. In retrospect, we suspect that a three-factor solution might even have been better: decoding skills, comprehension skills, and science knowledge.

A Multi-Year Perspective

Based on the results of our attempts to use, across several primary grades, indicators of (a) home environment, (b) pupil entry-level abilities, (c) home activities supportive of science learning, and (d) classroom process indicators to account for end-of-year science achievement, we have arrived at a few general conclusions:

1. In third grade, only the index of school activities brought home was significantly related to other variables in the model, having a (small) direct positive effect on end-of-year science achievement and a (smaller) negative effect on end-of-year verbal performance. Finally, in fourth grade, none of the home support indices had a relationship with other variables strong enough to be retained in the model.

   Collectively, these results suggest the diminishing role of activities in the home as influences on science achievement in the school. Yet the proportion of variance in end-of-year achievement that is accounted for in these models increases from grade to grade. Possible ways to account for this increase are considered next.

2. At the third-grade level, one classroom composite (Independent Study Activities), comprising five observed indicators (but primarily the frequency with which children worked independently at their desks), had a direct effect on end-of-year science achievement, but a small one. However, end-of-year verbal performance also had a significant direct effect on science achievement (β = .18) and was, in turn, affected by the other classroom composite (General Instructional Activities). Thus, verbal skill significantly influenced science performance at the end of third grade. By the end of fourth grade, though, the observed correlation between end-of-year verbal and science

performance was accounted for by other relationships in the model, so there was no direct effect of verbal performance on science. At the same time, though, the effect of beginning science performance was somewhat greater than in previous grades, and contributions from the classroom variables--individually or as indicators in a composite--were rather small (ß's ranging from -.11 to .11).

One interpretation of these various relationships is that teacher behaviors, as indicated by the classroom process indicators, had a greater effect in second grade than in later grades; that pupils whose reading skills (Verbal Performance) were better in third grade also attained higher levels of science knowledge, perhaps because of a greater emphasis on "reading to learn" at this grade level; and that this end-of-third-grade advantage in science achievement was then maintained through fourth grade. If this after-the-fact reasoning is correct, we would expect it to be supported by a multi-year structural model, and we might even predict that the fifth-grade results would be quite similar to those from fourth grade.

3. It is possible that we have masked some findings by pooling the students from these three districts for these analyses. The differences in sample sizes and teachers' behaviors between districts may have allowed effects in the largest districts to overpower effects in the two smaller districts. Future analyses will be done on a district by district basis. This will allow us to investigate science instruction in three different contexts.

Concluding Remarks

When we reported kindergarten through second-grade trends for science learning (Meyer et al., 1992) we finished on an optimistic note because the second-grade model showed positive effects from schooling. These effects were not realized at the third- and fourth-grade levels. There do appear to be two companion trends in these data, however.

First, in third and in fourth grade, children's entering ability has affected teachers' behaviors and activities. These sensitivities and responses were not apparent in the lower grades. Second, independent study has a small positive relationship to end-of-year science performance in third grade but a small end-of-year negative relationship to science performance in fourth grade.

The lack of relationships between the various classroom instructional constructs and end-of-year science performance illustrates most simply that the things that these teachers are doing simply are not affecting the children's science learning. The next logical question is, why? We can only speculate about the reasons for this.

Time allocated for science at the third- and fourth-grade levels falls far short of time generally allocated for subject areas deemed to be highest priorities in the lower grades. Children most often have literacy-related instruction for well over half an hour every day. The most successful classroom in the early elementary grades that we reported for these same children in reading (Meyer, Wardrop, & Hastings, 1990) averaged over 40 minutes each day--close to *twice* what these students received in science. Therefore, we suggest that one clear problem is that teachers simply have not allocated enough time to teach science. While simply increasing the time alone would not solve problems, it would at least create opportunities for more instruction.

We also observed teachers interacting with students at *low* rates in these science classrooms. For example, children averaged either below one turn each or slightly over a turn each for text-tied and review questions at these grade levels. Concept application interactions were almost nonexistent.

Instances of teachers manipulating objects and having students do the same were so infrequent that those data were not even part of the models. Similar problems exist for questions teachers asked to have children generate hypotheses and feedback teachers gave after students made errors. If children made mistakes, teachers seldom sustained interactions until they come up with correct answers. Taken together, it seems that these students were seldom questioned during instruction, *and* they very rarely were asked to apply the information they had received. In short, very little time was allocated to science instruction and little happened during the short periods they had. Contrary to popul. belief, we failed to see teachers relying heavily on textbooks. They were present, but they did not really dominate science instruction. Therefore, it is not as if teachers have abandoned hands-on activities in favor of text-based science instruction. They have, it appears virtually abandoned science instruction all together.

Our hypothesis was confirmed to the extent that we observed so little science instruction taking place in these classrooms that student performance in science continues to be a product of entering ability. In short, students who knew the most science when they started, knew the most when the school year ended.

Overall, it would seem that home influences diminish in these grades, schooling is of little or no influence, and therefore entering ability grows in its power to predict later performance. These findings, unfortunately, confirm Tifft's (1989) prediction that "A crisis looms in science" for the United States. In fact, for children in American schools, the crisis may have already arrived.

References

Abruscato, J., Fossaceca, J. W., Hassard, J., & Peck, D. (1980). *Holt elementary science*. New York: Holt, Rinehart & Winston.

Anderson, V., Bereiter, C., & Smart, D. (1977). *Catching on*. LaSalle: IL: Open Court Publishing Company.

Calfee, R. C., & Calfee, K. H. (1982). *Interactive reading assessment system (IRAS)*. Austin, TX: Southwest Educational Development Laboratory.

Educational Testing Service (1979). *Sequential test of educational progress*. Menlo Park, CA: Addison-Wesley.

Engelmann, S. E., and Hanner, S. (1972). *Distar reading III, reading to learn*. Chicago, IL: Science Research Associates.

Fort, D. C. (1990). From gifts to talents in science. *Phi Delta Kappan, 71*, 665-671.

Hastings, C. N., Meyer, L. A., Linn, R. L., & Wardrop, J. L. (in press). *Assessing science concepts and process acquisition for kindergarten, first-, second-, and third-grade children*. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Jastak, S., & Wilkinson, G. S. (1984). *Wide range achievement test*. Wilmington, DE: Jastak Associates.

Joreskög, K. G. (1978). Structural analysis of correlation and covariance matrices. *Psychometrika, 43*, 443-447.

Joreskög, K. G., & Sorbom, D. (1984). *LISREL VI user's guide*. Mooresville, IN: Scientific Software.

Mallinson, C. G., Mallinson, J. B., Smallwood, W. L., & Valentino, C. (1985). *Silver Burdett science*. Morristown, NJ: Silver Burdett.

Mechling, K. R., & Oliver, D. L. (1983). Activities, not textbooks: What research says about science programs. *Principal, 62*(4), 41-43.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1986). *Motion test*. Unpublished test. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1987a). *Living things test*. Unpublished instrument. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1987b). *Science passage test*. Unpublished instrument. Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Wardrop, J. L., & Hastings, C. N. (1990). *The development of reading ability in kindergarten* (Tech. Rep. No. 515). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Wardrop, J. L., & Hastings, C. N. (1992). *The development of science knowledge in kindergarten through second grade* (Tech. Rep. No. 549). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Open Court Publishing Company (1987). *Open Court test.* LaSalle, IL.

Sund, R. B., Adams, D. K., & Hackett, J. K. (1982). *Accent on science.* Columbus, OH: Charles E. Merrill.

Tifft, S. (1989, September 11). A crisis looms in science. *Time*, pp. 68-70.

Woodcock, R. W. (1973). *Woodcock reading mastery tests.* Circle Pines, MN: American Guidance Service.

Author Note

We gratefully acknowledge the support and cooperation from the school district personnel and children in the three districts participating in this work. We are also indebted to our field staff; Joy Blai , Barbara Boyer, Carol Chasnoff, Lorraine Crummey, Julie Custer, Amy De Hoogh, Kim Drone, Louise Dyer, Marlene Engberg, Tina Hunter, Brenda Ritzhaupt, Judy Ruzicka, and Mary Zwoyer. These persons completed classroom observations under the direction of Lorraine Crummey. We also appreciate their assistance analyzing the science textbooks and testing children.

Table 1

Means and Standard Deviations for all Measures of Student Ability

Instrument	Administered	$\bar{X}$	SD
WRAT, Level 1 Reading Subtest	fall, third grade	62.19	8.30
Woodcock Passage Comprehension Test	fall, third grade	38.94	11.63
IRAS: Average Passage Errors	fall, third grade	.11	.94
IRAS: Average Passage Rate	fall, third grade	.10	1.03
IRAS: Average Passage Self-Corrected Errors	fall, third grade	.03	.55
IRAS: Passage Comprehension Questions	fall, third grade	57.47	14.21
Circus: Listening Level D	fall, third grade	32.62	4.04
Circus: Think It Through Level D	fall, third grade	24.35	5.43
Science Passages Average Rate	spring, third grade	2.50	.51
Science Passages Average Decoding Errors	spring, third grade	23.96	20.54
Science Passages Average Self-Corrects	spring, third grade	2.07	.78
Science Passages Average Comp Score	spring, third grade	9.66	2.43
Open Court	spring, third grade	14.94	5.01
Living Things Test	spring, third grade	102.50	9.09
STEP: Science	spring, third grade	42.16	5.19
Motion Test	spring, third grade	13.23	2.66

Table 1 (Continued)

Instrument	Administered	$\bar{X}$	*SD*
WRAT, Level 1 Reading Subtest	fall, fourth grade	69.32	8.69
Woodcock Passage Comprehension Test	fall, fourth grade	47.16	11.00
STEP: Listening	fall, fourth grade	14.71	3.42
STEP: Reading	fall, fourth grade	42.82	8.63
Motion Test	spring, fourth grade	14.28	2.78
Open Court	spring, fourth grade	16.95	4.96
NSF Test	spring, fourth grade	59.83	9.60
Living Things Test	spring, fourth grade	56.67	8.18
Science Passages Average Background Knowledge	spring, fourth grade	4.20	1.13
Science Passages Average Rate	spring, fourth grade	4.24	1.85
Science Passages Average Decoding Errors	spring, fourth grade	18.63	19.73
Science Passages Average Self-Corrects	spring, fourth grade	3.65	3.03
Science Passages Average Comprehension Score	spring, fourth grade	9.32	2.38
STEP: Science	spring, fourth grade	30.95	10.48

Table 2

Means and Standard Deviations by District for Time Spent in Third Grade Science

District	N of Children	Time in Minutes/Observation	
		$\bar{X}$	*SD*
A	82	18.91	6.31
B	149	18.40	7.22
C	75	14.55	2.70

Table 3

Means and Standard Deviations by District for Interaction and Feedback Process Variables During [illegible]rd Grade Science Instruction

	Interactions									
	Procedural		Scriptal		Text-Tied		Gen. Review		Oral Rdg.	
District N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A 82	4.91	1.10	1.61	.68	.53	.45	.50	.75	.17	.18
B 149	7.34	3.30	.98	.54	.80	.71	.69	.79	.17	.29
C 75	5.70	1.80	.58	.26	*.82	.41	*1.08	.47	*.56	.31
	Feedback									
	Sustaining		Terminal		Confirming		Seg Ind. Wk		No Sci Dom Obs	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A	.14	.16	.16	.16	1.66	.71	16.61	1.64	3.34	.48
B	.27	.45	.13	.22	1.26	.73	*29.28	2.02	*4.15	1.25
C	.06	.10	.42	.17	1.21	.29	6.83	.00	3.00	.00

Table 4

Means and Standard Deviations by District for Time Spent in Fourth Grade Science Instruction

District	N of Children	Time in Minutes/Observation $\bar{X}$	*SD*
A	78	18.42	2.69
B	145	23.42	11.99
C	80	22.88	4.00

Table 5

Means and Standard Deviations by District for Interaction and Feedback Process Variables During Fourth Grade Science Instruction

		Interactions			
		Gen Review		Concept App	
District	*N*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
A	78	1.31	1.34	.00	.02
B	145	1.58	1.36	.68	1.02
C	80	1.31	1.08	.10	.10

		Feedback			
		Sustaining		Terminal	
District	*N*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
A	78	.11	.15	.07	.13
B	145	.40	.39	.19	.17
C	80	.37	.43	.94	.71

		Management							
		% on Task		Sweeps/Yr		Class Praise		Ind Critics	
	N	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
A	78	86.61	6.13	11.93	3.19	1.05	1.45	4.93	1.34
B	145	95.48	2.95	13.08	3.10	.56	.54	3.29	1.38
C	80	85.96	3.85	7.65	1.19	1.16	.18	3.92	.95

Table 6

Descriptive Statistics for Home Support Variables, Third and Fourth Grades, and All Schools

		District A		District B		District C		Total Group	
Year	Variable	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
3	PTCACTV	8.58	6.79	9.87	7.20	10.37	7.04	9.62	7.06
	EXPRWAD	48.15	10.41	47.54	10.82	50.87	11.89	48.39	10.97
	SCHACTV	12.12.	4.16	11.58	4.72	9.76	4.30	11.31	4.54
4	EXPWADL4	45.60	11.83	47.53	10.17	47.31	11.10	49.93	10.85
	PTCPACT4	11.58	6.28	11.97	7.50	13.25	6.96	12.21	7.06

Table 7

Total Effects on Endogenous Variables, Grade 3

TOTAL EFFECTS OF KSI ON ETA

	HOMBKGND
SCPERFB3	0.396
VBPERFB3	0.371
SCPERFE3	0.481
VBPERFE3	0.338
SCHACTV3	0.107
GENLNSTR	0.000
NSTRMGMT	0.000
INDSTUDY	0.112

TOTAL EFFECTS OF ETA ON ETA

INDSTUDY	SCPERFB3	VBPERFB3	SCPERFE3	VBPERFE3	SCHACTV3	GENLNSTR	NSTRMGMT	
SCPERFB3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
VBPERFB3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SCPERFE3	1.036	0.173	0.000	0.187	0.060	0.012	0.000	0.054
VBPERFE3	0.000	0.925	0.000	0.000	-0.052	0.064	0.000	0.000
SCHACTV3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GENLNSTR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NSTRMGMT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
INDSTUDY	0.284	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8

Total Effects on Endogenous Variables, Grade 4

TOTAL EFFECTS OF KSI ON ETA

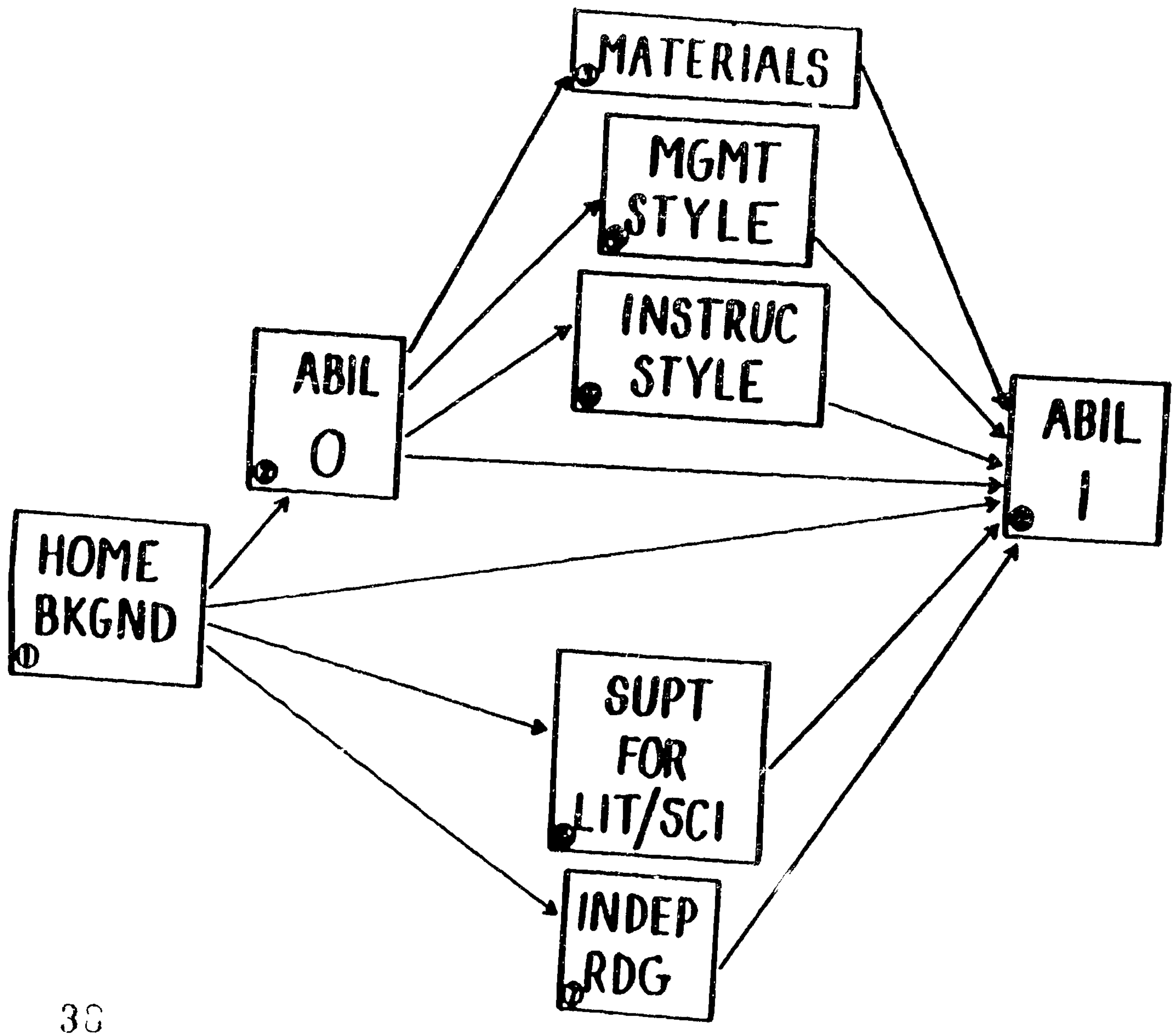
	HOMBKGND	FATHHRWK	MTHRHRWK
SCPERFB4	0.526	-0.154	0.000
VBPERFB4	0.211	0.000	0.081
NSTRATV	0.201	-0.059	0.000
INDSTATV	-0.036	0.000	-0.014
TERMFBK4	-0.040	0.000	-0.015
INDCRIT4	-0.134	0.036	-0.005
SCPERFE4	0.500	-0.146	0.000
VBPERFE4	0.201	0.004	0.083

TOTAL EFFECTS OF ETA ON ETA

	SCPERFB4	VBPERFB4	NSTRATV	INDSTATV	TERMFBK4	INDCRIT4	SCPERFE4	VBPERFE4
SCPERFB4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
VBPERFB4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
VSTRATV	0.382	0.000	0.000	0.000	0.000	0.000	0.000	0.000
INDSTATV	0.000	-0.171	0.000	0.000	0.000	0.000	0.000	0.000
TERMFBK4	0.000	-0.188	0.000	0.000	0.000	0.000	0.000	0.000
INDCRIT4	-0.231	-0.056	0.000	0.000	0.000	0.000	0.000	0.000
SCPERFE4	0.949	0.006	0.000	-0.107	0.099	-0.107	0.000	0.000
VBPERFE4	-0.027	1.020*	-0.072	0.000	0.000	0.000	0.000	0.000

*Presumed equal to 1.000 within rounding error.

Figure 1
Heuristic Model



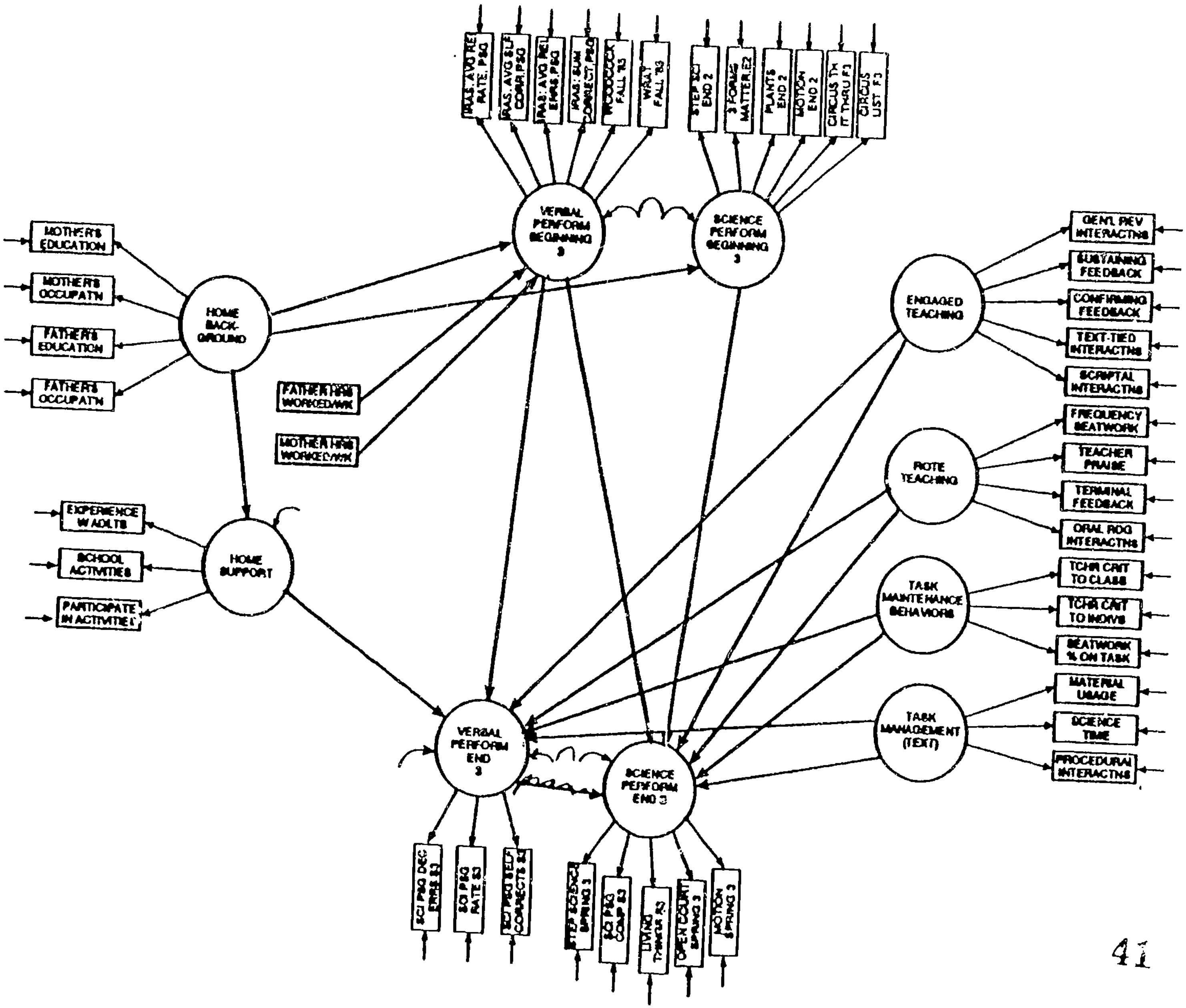


Figure 2
Initial Science Model for Grade 3, Cohort 1

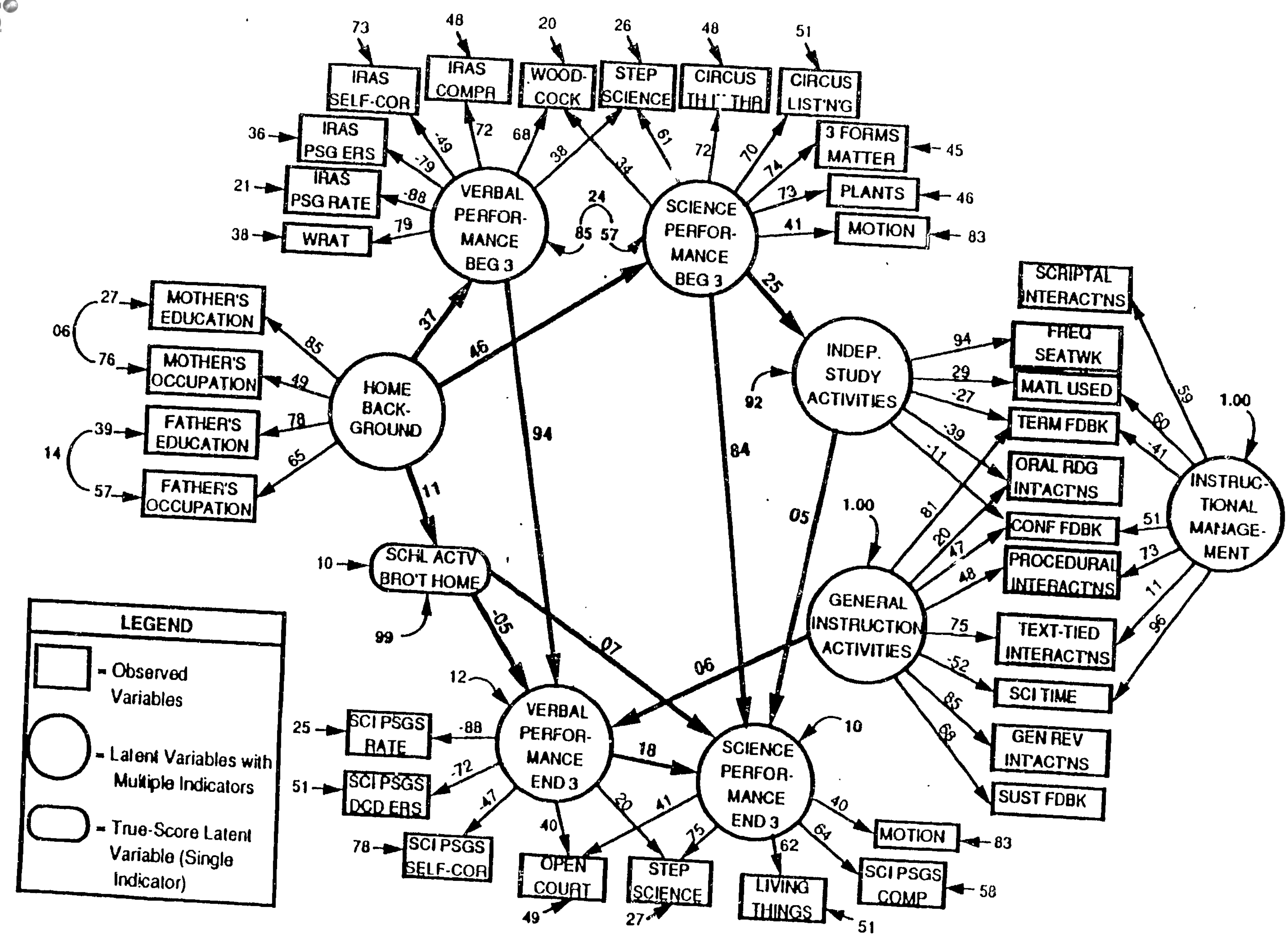


Figure 3
Final Structural Model for Science, Cohort 1, Grade 3

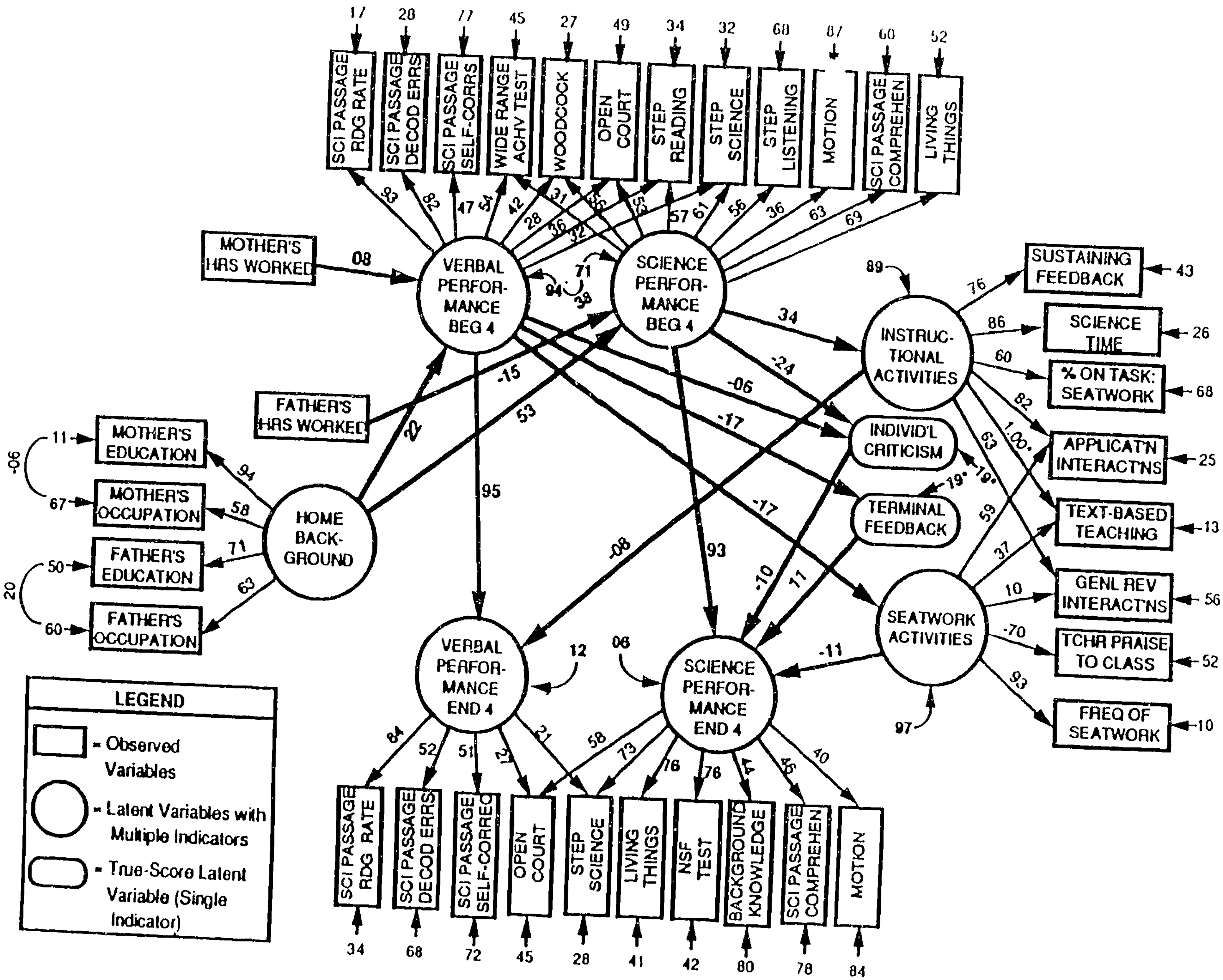


Figure 4
Final Structural Model for Science, Cohort 1, Grade 4