ED 301 148 IR 013 470

AUTHOR Mandinach, Ellen B.; Thorpe, Margaret E.
TITLE The Systems Thinking and Curriculum Innovation Project. Technical Report, Part 2.
INSTITUTION Educational Technology Center, Cambridge, MA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO ETC-TR88-12
PUB DATE Feb 88
CONTRACT 400-83-0041
NOTE 48p.; For Part 1 of this technical report, see IR 013 329.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests; *Cognitive Processes; *Cognitive Tests; Comparative Analysis; Computer Simulation; Computer Software; Correlation; Curriculum Development; *Intermode Differences; Mathematics Curriculum; Secondary Education; *Secondary School Science; Social Studies; *Systems Approach
IDENTIFIERS Brattleboro Union High School VT; *Systems Thinking and Curriculum Innovation Project

ABSTRACT

This is the second of two reports on the first year activities and results of the Systems Thinking and Curriculum Innovation Project (STACI), a two-year project which is examining the cognitive demands and consequences of using the STELLA (Structural Thinking Experimental Learning Laboratory with Animation) software to teach systems thinking, content knowledge, and problem solving. This report focuses on the performance of students at Brattleboro Union High School (Vermont) in physical science, biology, chemistry, and social studies on standardized and other tests; traditionally taught courses provided control. Four types of data are analyzed: (1) performance on achievement and ability tests and differences between students in traditional and systems classes within subjects; (2) performance on a reference battery of four cognitive tests and differences between subject classes; (3) performance on content-knowledge tests and the relationship between content and achievement tests; and (4) performance on systems thinking instruments, including the characteristics of the tests and the performance of students in different subject areas. Finally, interrelationships among all four types of data are examined and correlations are noted. The general effectiveness of each type of test is considered, and possible adjustments, changes, and revisions for the coming academic year are discussed. In addition, a new focus on self-regulated learning is detailed. Analyses of the data are presented in six tables within the text and five appended tables. (10 references) (EW)

THE SYSTEMS THINKING AND CURRICULUM INNOVATION PROJECT

Technical Report, Part 2
February 1988

Educational Technology Center
Harvard Graduate School of Education
337 Gutman Library Appian Way Cambridge MA 02138
(617) 495-9373

The Systems Thinking and C'urriculum Innovation Project

Technical Report, Part 2
February 1988

Prepared by

Ellen B. Mandinach
Margaret E. Thorpe

Educational Testing Service
Princeton, New Jersey

STACI Project

<u>Group Members</u>

Nancy L. Benton
Hugh F. Cline
Ellen B. Mandinach
Margaret E. Thorpe

Preparation of this report was supported in part by the Office of Educational Research and Improvement (Contract # OERI 400-83-0041). Opinions expressed herein are not necessarily shared by OERI and do not represent Office policy.

The Systems Thinking and Curriculum Innovation Project: Technical Report, Part 2

The present report, in conjunction with its companion (Mandinach & Thorpe, 1987), provide documentation of the first year activities and results from the Systems Thinking and Curriculum Innovation (STACI) Project. The first part of the report provides descriptions of systems thinking, Brattleboro Union High School (BUHS), the design, data collection, instrumentation, ancillary studies, and curriculum development. The present document discusses performance on achievement and ability tests, the reference battery of cognitive tests, content-knowledge tests, and systems thinking instruments.

At the outset of this document, it is critical to note that the data analyses reported here should be considered exploratory (see Mandinach, 1987 for complete rationale) due to several factors. Foremost, the data collection and subsequent analyses necessarily paralleled curriculum development. Throughout the academic year 1986-1987, the BUHS teachers developed their systems thinking curriculum modules with the intent to test them in a preliminary manner. The modules were implemented, tested, and have been revised for the 1987-1988 academic year. Given that the teachers conducted trial runs of their instructional materials, Educational Testing Service therefore developed instrumentation that could be flexibly revised, based on the results of the first year analyses.

Communication between BUHS and ETS facilitated revision of both curriculum and instrumentation/data collection. Results from the instrumentation enabled ETS to make changes in the tests and cognitive exercises as well as to help the teachers revise their curriculum modules. Conversely, feedback from the teachers assisted ETS in focusing, making more efficient, and modifying data collection procedures.

Thus, the analyses reported in the present document represent exploratory data collection that will be formalized in the STACI Project's second year. The analyses serve two primary functions. As noted above, they serve to stimulate revisions of curriculum development and data collection procedures. Second, the data provide general descriptive information about student performance on the cognitive reference battery, content knowledge tests, and measures of systems thinking. We are mindful that it would be premature to draw formal inferences about the impact of the systems treatment by comparing performance in the systems thinking and traditional courses. However, preliminary comparisons are discussed at this time about initial subject differences between the two treatment groups.

Results

Discussions of the results will focus on four types of data. First, ability and achievement test performance is reported. These data include the California Achievement Test (CAT) and the Advanced Progressive Matrices (Raven, 1958, 1962) which serve as measures of general ability. Other tests from the reference battery serve as measures of cognitive skills hypothesized to be related to systems thinking. Third, performance on content tests in general physical science, biology, and chemistry is documented. A discussion of the relationship of these tests to the systems thinking curricula also is provided. Finally, performance on the systems thinking instrument is reported. A complete description of the rationale for and content of this test is discussed. Results from all of the tests also are described in terms of revisions for future administrations.

The various analyses reported here require examination of different parts of the sample. Consequently, the sample sizes will differ in accord with the measures used in the analyses. Two caveats should be noted. First, the CAT scores for the freshmen class were not available. Because freshmen were taking general physical science (GPS), analyses that focus on the achievement test do not include the GPS classes. Second, students in the War and Revolution seminar are included in some of the analyses to provide an upper boundary of performance (particularly for the systems thinking instrument). However, they are not included in the content test analyses, which focus only on the science classes.

Ability and Achievement Measures

Two measures of ability were used to gauge students' general intellective functioning. First, students' most recent standardized achievement test scores were used to assess general crystallized ability (Cattell, 1971). Crystallized ability refers to previously constructed assemblies of performance processes retreived as a system and applied anew in familiar instructional situations (Snow, 1980, 1982). This construct reflects long-term accumulation and organization of knowledge and skills. The reading, language, and mathematics subscales and total test score from the CAT were extracted from students' school records. To enhance interpretability, national percentiles will be reported here.

Second, Parts I and II of the Advanced Progressive Matrices (APM) were used to assess general fluid ability. Fluid ability refers to new assemblies adapted to new and unfamiliar situations and novel problems by reassembly of available performance processes (Snow, 1980, 1982). For the 1986-1987 administration of the matrices, the six odd items from Part I and the first nine odd items from Part II were given to the students as part of the cognitive reference battery. The reliability for the instrument, using the Spearman-Brown formula for unequal lengths, was calculated at r=.80.

On average, students performed well on the APM (M = 11.83, S.D. = 2.81). Four of the items seemingly discriminated among levels of performance (see Appendix A). A distinct developmental progression was noted, $F(2, 362) = 21.83$, $p < .001$ (n.b., GPS students generally were freshmen, biology mostly sophomores, and chemistry juniors or seniors). Students in the War and Revolution seminar answered nearly all of the test correctly (M = 14.70), whereas those in GPS scored the lowest (M = 10.48). The biology (M = 11.81) and chemistry (M = 12.83) students scored somewhere in between the extremes. Furthermore, students in the traditional classes (M = 12.08) performed slightly better than those in the systems classes (M = 11.43), $F(1, 362) = 5.95$, $p < .05$.

Data from the CAT were available for most of the biology and chemistry students, but for only six GPS students. Consequently, the GPS students have been omitted from comparative analyses across classes. The BUHS students averaged in the 78.49 percentile (S.D. = 18.03) in reading, 73.84 (S.D. = 20.73) in language, 76.20 (S.D. = 18.86) in mathematics, and 77.37 (S.D. = 18.88) on the total test.

Appendix B presents the results of the achievement tests for the systems thinking and traditional classes in biology, chemistry, and War and Revolution. The biology students scored lower than those in chemistry on the CAT. The data also indicated that the students in War and Revolution were outliers, compared to others in the sample as well as nationally.

Statistical analyses were performed to determine if there were pre-existing differences in achievement test scores between the treatment and systems classes within subjects. No significant differences were found between treatments for chemistry on any of the test scores (reading $F(1, 113) = .014$, ns; language, $F(1, 113) = .48$, ns; mathematics, $F(1, 113) = .01$, ns; and total test $F(1, 113) = .02$, ns). However, differences were noted across the test in the biology classes. The systems classes scored lower than the traditional students on reading, $F(2, 142) = 6.48$, $p < .01$; language, $F(2, 142) = 3.54$, $p < .05$; mathematics, $F(2, 142) = 6.65$, $p < .01$; and total test $F(2, 142) = 4.25$, $p < .05$.

The above results indicated that older students performed better on the measures of ability and achievement. This developmental trend should be noted but is less of a concern, given that most analyses this year and next will focus on within subject area differences. However, the analyses also indicated differences between treatment conditions on measures of ability. The two chemistry groups were comparable on achievement test scores, whereas the traditional biology classes scored significantly higher than did the systems classes. In contrast, the biology classes were roughly comparable on the APM, whereas both the traditional GPS and chemistry classes scored higher on this measure. These initial treatment differences will be accounted for in subsequent analyses.

Reference Battery

Four other tests were administered with the APM in a reference battery. The intent was to assess students' performance on skills related to systems thinking. The four tests were the Figural Analogies (FA), Diagramming Relationships (DR) and Letter Sets (LS) (Ekstrom, French, and Harman, 1976), and Deductive Reasoning (DRT). The latter test will be dropped from the present discussion for two reasons. First, one teacher was unable to administer the test to her class. Second, results and feedback from the teachers indicated that the students did not understand the task nor could the teachers administer it in such a way to explain the problem. Thus, the test will be eliminated from the current analyses. A revised task has been developed and will be administered during the 1987-1988 academic year.

Figural Analogies. FA consisted of 10 items presented in a four-alternative, multiple-choice format. A figural analogy item consisted of a typical analogy item (A:B::C:D), but used geometric figures, such as triangles, squares, circles, etc., rather than words. Mean performance on the items ranged from .06 to .96 (p values were greater than .70 for all but three items), with a split-half reliability of $r = .53$. Two items appeared to present particular difficulty (see Appendix C). Both items represented geometric addition. Item 4 ($p = .04$) required students to infer that a triangle is to a square as a square is to a pentagon. That is, 3:4 (3+1)::4:5 (4+1). Similarly, Item 8 ($p = .35$) required students to infer, in terms of shaded areas of geometric figures, that 1/4:1/2::1/2:1.

Performance differences were noted on the FA. The systems ($M = 7.06$) and traditional ($M = 7.01$) groups performed comparably, $F(1, 362) = .06$, ns, whereas a developmental trend was found across subject areas, $F(2, 362) = 10.96$, $p < .001$. Students in chemistry ($M = 7.53$) scored the highest, followed by biology ($M = 6.99$), then GPS ($M = 6.51$).

Diagramming Relationships. DR consisted of 15 items presented in a five-alternative, multiple-choice format. The purpose of the test was to see if students were able to discern and diagram relationships among groups of things. For example, a sample problem asks students to determine the relationships among birds, pets, and trees. The correct response would indicate that trees are neither pets nor birds, but some birds are pets. Thus, there should be an intersection between pets and birds, but not one between trees and pets or birds.

Mean performance on the items ranged from .28 to .92, with an average total score of 8.40 (S.D. = 3.24). There were 11 items on which the p values were under .70 (see Appendix D). The task obviously was difficult and unlike any the students had encountered previously. The split-half reliability (Spearman-Brown for unequal parts) was $r = .74$.

A similar pattern of performance differences was noted on the DR. The systems ($M = 8.49$) and traditional ($M = 8.15$) groups performed comparably, $F(1, 362) = 1.00$, ns, whereas a developmental trend was found across subject areas, $F(2, 362) = 14.59$, $p < .001$. Students in chemistry ($M = 9.31$) scored the highest, followed by biology ($M = 8.41$), then GPS ($M = 7.03$).

Letter Sets. LS also consisted of 15 items presented in a five-alternative, multiple-choice format. Each item contained five sets of of letters in groups of four. A common rule linked four of the five letter sets. LS required identification of a rule that made four sets of letters similar in some way, and one set different from the others. For example, NOPQ DEFL ABCD HIJK UVWX, is one of the sample items. The rule here is four letters in alphabetical sequence. The second set violates that rule, and thus is the correct response.

Mean performance on the items ranged from .10 to .96, with an average total score of 10.34 ($S.D. = 3.02$). Performance on this task was more consistent. However, the p values indicate that students may have had difficulty finishing the task within the time constraints. Performance on Items 11, 13, 14, and 15 reflect this possibility (see Appendix D). The split-half reliability (Spearman-Brown for unequal parts) was $r = .80$.

Again, a similar pattern of performance differences was noted on the LS. The systems ($\underline{M}$ = 10.10) and traditional ($\underline{M}$ = 10.47) groups performed comparably, $\underline{F}(1, 362) = 1.62$, ns, whereas a developmental trend was found across subject areas, $\underline{F}(2, 362) = 11.90$, $\underline{p} < .001$. Students in chemistry ($\underline{M}$ = 10.76) and biology ($\underline{M}$ = 10.76) scored the highest, followed by GPS ($\underline{M}$ = 9.09).

<u>Relationship Among the Ability and Reference Battery Measures</u>

Table 1 presents the intercorrelations among the CAT, APM, and the reference battery measures. The correlations among the reference battery measures were found to be moderate, but significant. Similar correlations were found between the reference battery and the achievement test scores. Intercorrelations between the subscales of the CAT were extremely strong.

<u>Content Tests</u>

Pretests (in September) and posttests (in June) in the science courses were given to assess initial and ending levels of content-specific knowledge. It was expected that students would have difficulty on the pretests, given that they were administered before any formal instruction in science had been undertaken.

The posttests served two purposes. First, they provided information about students' levels of knowledge at the end of the courses. The tests also were intended to provide means of comparisons be'ween systems and traditional classes on topics that had been taught from a systems perspective. However, the incidences of overlap between test and systems concept coverage was not sufficient from which to draw any conclusions. This information has been given to the teachers with whom ETS will work to make more sensitive tests for next year.

The two GPS tests each contained 40 multiple-choice items, with a maximum possible score of 40 points. Comparisons between the systems (M = 16.54) and traditional (M = 16.31) classes on the pretest indicated no initial differences, $F(1, 91) = .08$, ns. A treatment difference was noted on the posttest on which the systems classes (M = 19.74) performed better than the traditional students (M = 17.62), $F(1, 95) = 4.37$, $p < .05$.

The biology tests contained 44 multiple-choice items. No initial differences were found either across the three teachers, $F(2, 143) = 2.12$, ns, or between treatment groups, $F(1, 143) = .84$, ns, on the pretest. Average performance was 14.61 items correct. However, a significant difference was noted on the posttest. Students in the two traditional teachers' classes (M = 22.42 and 22.83) were roughly comparable. The systems teacher's classes did not perform as well (M = 20.31), $F(2, 144) = 3.91$, $p < .05$. Thus, the comparisons between treatment conditions indicated that students in the traditional classes (M = 22.62 versus M = 20.31), performed significantly better on the biology posttest, $F(1, 144) = 7.74$, $p < .01$. Further examination at the class level indicated that all four of the systems classes performed lower than the traditional classes. However, one class in particular drove the systems treatment average (M = 18.9). The average for the other three classes was 20.9. An additional explanation for this finding is discussed below in terms of the impact of ability on the performance differences.

The chemistry tests contained only 20 multiple-choice items. The pretest proved to be extremely difficult as several students failed to get any item correct (even with guessing, one would expect a few chance correct answers). No treatment differences were found on the pretest, with equivalent perform by the systems (M = 4.47) and traditional (M = 4.23) classes, $F(1, 116) = .29$, ns. A significant treatment difference was found on the posttest, $F(1, 116) = 8.11$, $p < .01$. The systems classes (M = 10.25) outperformed the traditional students (M = 8.64) on posttest content knowledge.

Relationship Between Content and Achievement Tests

Given that the CAT scores served as measures of academic achievement, we examined the relationship between CAT and content test performance. Positive correlations were found in both biology and chemistry classes. Biology posttest performance was related to reading, $r(134) = .48$, $p < .001$; language, $r(136) = .48$, $p < .001$; mathematics, $r(135) = .47$, $p < .001$; and total test score, $r(130) = .54$, $p < .001$. Chemistry posttest performance was related to reading, $r(106) = .47$, $p < .001$; language, $r(106) = .48$, $p < .001$; mathematics, $r(105) = .52$, $p < .001$; and total test score, $r(104) = .58$, $p < .001$. Thus, the more able students, as measured by CAT scores, also performed better on the content knowledge posttests.

Systems Thinking Instrument (STI)

The remainder of this document will focus on the primary instrument by which knowledge acquisition in systems thinking was assessed. The rationale for and content of the instrument will be described first. Results from the first year administration then will be presented.

Test characteristics. The STI consisted of 76 items that were intended to assess a range of skills thought to underlie systems thinking. A rational task analysis yielded 10 skills which then were made into subscales of varying lengths. Table 2 presents the alpha reliabilities for those scales. The 76 items on the total test yieldea an alpha of .95, indicating that the test was extremely consistent across items and subscales.

The understanding of basic graphing concepts (e.g., labeling and scaling axes, coordinates) comprised the first subscale. The seven graphing items yielded and alpha of .51. This wa the least internally consistent scale in the instrument. Two other subscales focused on graphing skills. A first required the interpretation of graphs. That is, students were asked to interpret a graph and provide a verbal description. The seven items on this scale yielded a reliability of .72. The second subscale required translation. That is, students were asked to take a verbal description of a problem and translate it into a graphical representation. There were seven such items (= .68).

Two other scales focused on mathematical skills that relate to systems thinking. First, three mathematical equation items were constructed, which yielded a reliability of .67. It was assumed that in order for students to become adept at using the STELLA software, they should have a basic understanding of how to solve simple equations. The second scale focused on students' understanding of graphical functions. Students were asked to write functions, find slopes, and graph functions. The five items yielded an alpha of .87.

Three scales were designed to assess knowledge of concepts that are critical in systems thinking. The first focused on variables (four items, = .92). These items required students to differentiate between dependent and independent variables, place them on the appropriate axes, and interpret the graphs' meaning. Three items then targeted the notion of causality (= .92). Here students were asked to complete a causal relationship and a causal diagram. Defining causality leads directly into the concept of looping. Five items (= .86) required that students either interpret or construct a causal loop diagram.

The final two scales measured skills and knowledge unique to systems thinking, STELLA, and modeling. The first of these subscales consisted of simple identification items. One set of identification questions required students to determine if a variable was a stock (level) or a flow (rate), then define its unit of measure. A second set of identification items asked students to identify parts of a structural diagram (e.g., stock, connector, flow, converter). The 25 items yielded an alpha of .92. The final subscale focused on the construction and interpretation of systems models of varying complexity. There were 10 items in the scale which yielded an alpha of .87. At one end of the continuum, students were asked to take a simple model and identify how certain variables affect other variables. At the other end of the continuum, students were asked to take a verbal description of a problem, construct a model, and then interpret that model.

There was a fundamental rationale for constructing such a diverse test. First, we wanted to examine how the items performed and how they related to what the students had encountered in class. That is, we needed to construct an instrument that would be both reliable and valid with respect to their classroom experiences with systems thinking. Performance on such an test would inform ETS and the BUHS teachers in terms of students' learning outcomes and cognitive processing, as well as for revisions of future instrumentation.

Second, and perhaps most importantly, we wanted the instrument to be reflective of and sensitive to differences in the systems curricula across the subject areas. As documented in the previous report (Mandinach & Thorpe, 1987), each of the teachers introduced systems concepts to their classes differently. That is, they focused on particular concepts of varying difficulty and integrated them into their courses with different degrees of intensity. For example, the GPS classes stressed understanding of measurement concepts, but did not spend much time on modeling. The biology classes did construct some models, whereas the chemistry classes were given models and asked to modify them. The War and Revolution seminar achieved a high level of sophistication with systems thinking, modeling, and STELLA. Thus, the instrument should yield ranges of performance within the courses around concepts that were either stressed or briefly described. The STI was constructed to provide information about these ranges that would be reflective of the curricular differences. Analyses of performance in the project's second year will focus on the range of differences related to instructional emphases.

Interrelationships among the subscales. Table 3 presents the intercorrelations among the STI's 10 subscales. All correlations were found to be significant at the $p < .001$ level. Several relationships should be noted. First, the correlations among the graphing scales were quite strong. Correspondingly, interpretation and translation of graphs were related strongly and positively to the understanding of looping and causality. Performance on the two mathematically oriented scales (equations and graphing) were strongly related. Understanding the basic concepts of systems thinking (the Systems Identification scale) was strongly related to performance on the Graphing Translation, Causality, and Loops scales.

Interestingly, the Systems Identifications items were least related to performance on the Variables scale, which showed the lowest correlations with the other scales. As expected, the two systems scales showed the strongest relationship of all the intercorrelations on the instrument, $r(177) = .72$, $p < .001$. Also as hypothesized, many of the subskills that underlie model construction were related to the Systems Interpretation scale. Performance on both of the complex graphing scales (Interpretation, $r(177) = .51$, $p < .001$ and Translation, $r(177) = .56$, $p < .001$), Causality, $r(177) = .55$, $p < .001$, Loops, $r(177) = .61$, $p < .001$, and Mathematical Graphing, $r(177) = .61$, $p < .001$ correlated with students' ability to construct, interpret, and manipulate models.

Test performance. Item-level performance data are presented in Appendix F. The appendix is broken down by subject area in order to examine the progression of skills and knowledge across courses. Table 4 presents the descriptive statistics for the 10 subscales. These data are further broken down by course in Table 5. Although these tables present the raw scores, subscale- and item-level data discussed in the text are reported in terms of percent correct in order to aid interpretability.

On all three of the graphing subscales (Graphing, Graphing Interpretation, and Graphing Translation) clear developmental progressions across courses were noted. Students in GPS generally scored the lowest, followed by biology, then chemistry. A ceiling effect was found for the War and Revolution students. They obtained a perfect score on the Graphing subscale and nearly perfect scores on Interpretation and Translation. In contrast, the GPS students obtained only 57, 35 and 34 percent correct on the scales, respectively.

More indepth analyses indicated particular problem areas experienced by the students. On the Graphing scale, the GPS students did not understand the (x, y) convention of defining points within a coordinate system. This deficit led to poor performance on items that required graphing points and defining axes. It also led to confusions when they were asked to translate or interpret graphical problems.

Other weaknesses of the GPS students were noted on the Mathematical Equations and Mathematical Graphing scales. On the latter scale, these students showed no understanding of functions. They were unable to graph functions, determine slopes, or write a function when given a graphical representation. It is possible that freshmen-level mathematics courses had not covered these concepts. The GPS students achieved only 8 percent correct on this subscale, in contrast to 30 percent for biology, 53 percent for chemistry, and 100 percent for War and Revolution. The GPS students performed only slightly better on the Mathematical Equations scale ($\underline{M}$ = .17). They experienced substantial difficulty in solving simple mathematical equations. In contrast, the biology students achieved 67 percent correct, chemistry had 76 percent, and War and Revolution gained 98 percent correct.

A different pattern of performance was noted on the Variable subscale. Here, the GPS students ($\underline{M}$ = .41) outperformed those in biology ($\underline{M}$ = .31). Chemistry students achieved 58 percent correct, whereas War and Revolution obtained 86 percent. Careful examination of the biology students' responses indicated that they did not understand the difference between independent and dependent variables nor could they provide a rationale for how they defined the sets of variables. They also did not know how to graph the variables (i.e., on which axes were the independent variables and dependent variables to be placed).

Yet on the two scales more directly related to variation and causality (Causality and Loop), the developmental progression across classes again appeared. The items in the Causality scale required students to complete a causal statement (e.g., Amount of studying causes ______.), then make the causal loop. GPS students could complete the statements, but had particular problems with the simple causal loops, obtaining only 31 percent correct. This is in contrast to 58 percent for biology, 65 percent for chemistry, and 98 percent for War and Revolution. Completion of more complex causal loops were required within the Loop scale. Performance on this scale was slightly lower but analogous to the Causality scale (M_{GPS} = .27; $M_{Biology}$ = .57; $M_{Chemistry}$ = .62; and $M_{W\&R}$ = .97). The GPS students experienced particular difficulty in translating a simple verbal description into a loop diagram. These results reflect the small amount of exposure the GPS students were given to loops.

On the more elementary scale that tested basic knowledge of systems thinking and STELLA, the Systems Indentifications scale, there were several notable trends. First, again there was a ceiling effect for the War and Revolution class (M = .94), particularly on the items that required identification of parts of a model (e.g., stocks, flows, connectors, converters). The GPS students (M = .42) had some trouble with the units of measure items as well as those that dealt with the parts of a model. Overall performance was roughly equivalent for the biology (M = .55) and chemistry (M = .59) classes. However, closer examination indicated that the chemistry students were more adept at the units of measure and variable items, whereas the biology students performed better on the structural modeling identifications. These performance trends were reflective of curricular emphases (time with and without STELLA) in the given courses.

A similar performance pattern was found for the Systems Thinking scale, where the focus was on modeling and interpretation of models. Given that the GPS students had little exposure to these complex concepts, they achieved only 19 percent correct. They at least made some attempt to interpret the models given to them, but had slightly more difficulty constructing their own models. The biology (M = .33) and chemistry (M = .35) classes again performed similarly, whereas the War and Revolution seminar did extremely well (M = .86). These trends also were expected, given the structure of courses and how systems thinking was integrated into them.

What is striking about the Systems Thinking scale is that every item on the subscale exhibited the same performance trend across courses (see Appendix F). The scale consisted of Items 35A, 35B, 35C, 36A, 36B, 37, 38A, 38B, 39A, and 39B. These items emphasized various forms of representations (i.e., structural diagrams, verbal descriptions of problems or models) and required students to apply their knowledge of systems thinking in the form of model interpretation or construction. In all cases, the War and Revolution seminar either achieved or approached a perfect score. These scores were well above any of the science classes. The GPS students achieved the lowest scores, whereas the biology and chemistry classes performed similarly and between the two extremes.

A more indepth examination of performance on the STI focused on Items 39a, 39b, and 39c, which were constructed to be reflective of the ultimate goals of the curriculum innovation in systems thinking. The three parts of the problem required the use of several higher-order thinking skills that were not only central to systems thinking, but also to general intellective functioning. The item required students to read, understand, and interpret in various ways a short word problem. Students were asked to consider the problem, then use analytic skills and hypothetical reasoning. They then translated the verbal description into a model that reflected understanding of the problem. The central concepts here were understanding variables, relationships, interactions, and being able to take the verbal description and represent it visually in the form of a model. A final requirement of the item was to have students take a small part of the problem, isolate particular variables within the model, then form a hypothetical relationship, based on altered parameters.

Performance on the three items (see Appendix F) indicated that, indeed, these are difficult skills and concepts tc apply. Item 39a, the simplest of the three, required fairly straightforward translation from the verbal problem description. The War and Revolution students had little difficulty here, scoring 94 percent correct. However, the science classes (M_{GPS} = .17; $M_{Biology}$ = .21; $M_{Chemistry}$ = .26) experienced a good deal of difficulty. There was an opportunity to generate several possible solutions to the problem. Yet many students were able to generate only a few. Responses on Item 39a indicated that students were unable or unwilling to provide multiple answers. This finding is not totally surprising, given that most tests they encounter contain items that have only one right answer. The notion that there may be infinite possible solutions perhaps was problematic. However, this was an underlying, but not explicitly stated, theme within the systems thinking curricula.

Item 39b was the problem most directly related to high-level systems modeling. Given its complexity, a wide range of performance, reflective of the teachers' emphases (or lack thereof) on modeling, was expected. Performance by the GPS classes (*M* = .06) reflected the small amount of time devoted to actual modeling. In contrast, the biology (*M* = .13) and chemistry (*M* = .15) classes spent some time on modeling, particularly the former. Whereas the chemistry classes focused on how changing parameters within existing models affects the systems, the biology classes emphasized model construction. Most of the students were at least able to construct the basic elements of the model. However, they failed to build-in many of the complex features. In contrast, most of the War and Revolution students (*M* = .83) were able to complete the entire model, reflective of the seminar's emphases on model building and problem solving. Only one of the seven students had difficulty with the model.

The intent of Item 39c (part of the Graphing Translation subscale) was to extend even the War and Revolution students beyond the information given in the basic problem. We wanted to see if the students were able to make the intuitive step beyond the explicitly stated data and hypothesize about potential interactions within the system. The War and Revolution class achieved 72 percent correct, with a few students completing the problem. Others in the seminar had difficulty with the problem. Many students in the science classes did not even make an attempt to solve the problem. However, some students in each of the classes managed to make progress on it (M_{GPS} = .18; $M_{Biology}$ = .18; $M_{Chemistry}$ = .23).

Relationships Among Ability, Achievement, Content, and Systems Thinking Tests

Analyses were conducted to see if the measure of general ability, the APM, affected the pattern of correlations among content and systems thinking test performance, within subject area. Table 6 presents the first-order correlations for the chemistry classes. As expected, the pre- and posttests of chemistry were related, $r(54) = .33$, $p < .01$. Posttest performance also was related to the systems thinking instrument, $r(54) = .35$, $p < .01$. However, when partialling out the APM, no change in correlations was noted (see Table 6). A different pattern emerged in the biology classes. The relationship between the biology pretest and performance on the systems thinking instrument was quite strong, $r(65) = .56$, $p < .001$. This magnitude of the correlation increased when APM scores were partialled out, $r(64) = .62$, $p < .001$. Clearly, general ability was a factor, to some degree, in the biology classes.

A third pattern was noted in the GPS classes, in which performance on the content pretest was not related to the systems instrument, $r(34) = .25$, ns. Yet the systems test was related to both the GPS posttest, $r(34) = .43$, $p < .01$, and to the APM, $r(34) = .47$, $p < .01$. When the APM was partialled out, the relationship between GPS and systems performance decreased, $r(33) = .32$, $p < .05$.

The relationship between the STI and CAT also were examined. These results are reported in Table 7. Achievement, as measured by the CAT, was related to performance on the STI. Not surprisingly, the reading subscale was most strongly related to the Graphing Interpretation ($r(119) = .52$, $p < .001$) and the Graphing Translation ($r(119) = .51$, $p < .001$) subscales. These two scales required the most use of verbal skills. A similar trend was found for the language scale (Graphing Interpretation, $r(120) = .53$, $p < .001$; Graphing Translation, $r(120) = .51$, $p < .001$). Performance on both the reading ($r(119) = .53$, $p < .001$) and language $r(122) = .56$, $p < .001$) was strongly related to the total STI score. Both the mathematics-oriented subscales (Equations, $r(121) = .45$, $p < .001$, and Graphing, $r(121) = .53$, $p < .001$) showed strong, positive correlations with CAT mathematics performance. Additionally, both Graphing Interpretation ($r(121) = .47$, $p < .001$) and Graphing Translation ($r(121) = .50$, $p < .001$) as well as total STI score ($r(121) = .52$, $p < .001$) were related to the CAT mathematics scale. Most of the STI subscales showed strong correlations with the CAT total score. Particularly strong were Graphing Interpretations, Graphing Translation, and Mathematical Graphing. The two total test scores were the most strongly related ($r(117) = .60$, $p < .001$) of all the correlations. Interestingly, the weakest relationship was between Systems Identifications across all of the CAT subscales.

Implications for Future Work

As stated earlier, the results reported here are preliminary in nature and are being used to provide information to researchers and teachers about adjustments, changes, and revisions for the 1987-1988 academic year. Several conclusions can be drawn from these results. First, the Deductive Reasoning Test, in its initial administration, did not yield useful data. This was in part due to task difficult and the teachers' inability to provide appropriate guidance to the students. The task has been revised and will be administered with the next reference battery.

Performance of the other reference battery measures was satisfactory. We will administer parallel forms this year to gain a second estimate of intellective functioning on the targeted skills as well as a measure of test reliability.

Although the results from the systems thinking instrument were highly reliable and satisfactory from a psychometric perspective, revisions were undertaken based on several pieces of feedback. The teachers requested a shorter and more focused instrument that would not discourage the GPS students. Experts at ETC provided content-related feedback that led to certain item-specific revisions. Finally, ETS revised several items to make them more interesting and relevant for the students, without changing the basic structure of the questions.

The results indicated a critical need for a content test that would measure the effects of the systems thinking curricula. As noted above, the science content tests failed to provide the requisite information for two reasons. First, there were too few items that reflected the systems modules. Second, the questions were not sufficiently demanding from a cognitive perspective. That is, they focused on factual or declarative knowledge rather than higher-order, procedural knowledge. Thus, ETS only could compare performance differences in content knowledge without being able to make any causal links to the curricula.

ETS is working with the BUHS teachers on two types of exercises that will assess curricular impact. First, the systems and traditional teachers will work together to develop content tests for particular modules that will be taught with systems thinking. That is, once the systems teacher identifies a topical area in which systems thinking will be used, both the systems and traditional teachers will construct a common examination to measure knowledge acquisition on that topic. A second type of measure will focus on more cognitively engaging problems. The teachers will select a production task that requires higher-order problem solving skills (rather than multiple-choice, factually based questions) on a topic that will be taught using the systems thinking approach. Both systems and traditional teachers will administer the problem to their students to compare more complex skills that have been affected by the systems thinking curriculum.

Another focus for the coming year will be to examine self-regulated learning. Briefly, self-regulation is a psychological construct that is procedural in nature and applicable across domains. Self-regulation refers to a student's active acquisition and transformation of instructional material (Corno & Mandinach, 1983). Self-regulation processes include discriminating relevant from irrelevant stimuli, connecting new information to existing knowledge structures, planning performance routines, alertness, and monitoring performance. These content-free skills are essential for cognitive processing, problem solving, and effective learning. Thus, they are a primary focus of the proposed research. These constructs have been described in more detail in previous documents submitted to ETC (see Mandinach, 1986, 1987).

Two procedures will be implemented to study self-regulated learning. First, building on items already in the STI, we can examine students' ability to use self-regulated learning processes such as monitoring, connecting, and alertness. The STI contains several multiple-part items (e.g., Items 28A-D and 29A-C) that approach one problem from different perspectives. For example, a student might be required to interpret a graph, translate a verbal description into a graph, or construct a causal loop diagram or model from either the verbal description or a graph. Each part of the item requires a very different representation of the problem. Some students may see the connections among the representations, whereas others may fail to see the interconnections. Self-regulation can be measured by examining how students work through these interconnected problems. Are they alert to the interconnections? Can they connect together the relevant parts? Can they monitor their performance across the interconnected parts, as different responses are required?

A second examination of self-regulated learning will occur in case study format. ETS will select a small subsample of students who exhibited interesting patterns of self-regulation on last year's STI. These students will be examined in case study format through the use of special tasks designed to measure self-regulation. We will observe and interview students, and examine their performance on written exercises.

Entering the second year of the STACI Project, we will be able to trace the impact of course taking patterns as students move between systems and traditional courses. Table 8 presents the dispersion of 1986-1987 students into the 1987-1988 academic year. There are more traditional than systems students in the second year's science courses who participated in the project's first year. From systems GPS, 14 students are taking systems biology and 19 taking traditional biology. From traditional GPS, 16 students are taking systems biology and 23 taking traditional biology. From systems biology, 21 students are taking systems chemistry and 34 taking traditional chemistry. There are 20 students from traditional biology now in systems chemistry and 33 in traditional chemistry. An equal number of systems and traditional chemistry students are enrolled in systems physics this year. Thus there will be sufficient opportunity to monitor progress and trace transfer as students are enrolled in the sequence of science courses.

Finally, the results will enable us to examine the effect of the systems thinking curriculum innovation on the teachers and their teaching activities. Most of the STACI Project results have focused on student learning outcomes. Yet, less readily quantifiable effects have influenced the teachers and their instructional procedures. ETS will attempt to examine these effects through interviews and the organizational case study.

References

Cattell, R. B. (1971). Abilities: Their structure, growth, and action. Boston: Houghton Mifflin.

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. Educational Psychologist, 18, 88-108.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). Kit of factor-referenced cognitive tests. Princeton, N.J.: Educational Testing Service.

Mandinach, E. B. (1986). Simulation software: A cognitive analysis and classroom study (Preliminary proposal submitted to the Edcuational Technology Center). Princeton, N. J.: Educational Testing Service.

Mandinach, E. B. (1987). Systems thinking and curriculum innovation (STACI) project. Proposal submitted to the Educational Technology Center and the Office of Educational Research and Improvement. Princeton, N.J. Educational Testing Service.

Mandinach, E. B., & Thorpe, M. E. (1987). The systems thinking and curriculum innovation project: Technical report, part 1 (TR87-6). Cambridge: Harvard University, Educational Technology Center.

Raven, J. C. (1958). Advanced progressive matrices (Set I). New York: Psychological Corporation.

Raven, J. C. (1962). Advanced progressive matrices (Set II). New York: Psychological Corporation.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P-A. Federico, & W. E. Montague (Eds.), Aptitude, learning, and instruction (Vol. 1). Hillsdale, N. J.: Erlbaum.

Snow, R. E. (1982). Training of intellectual aptitude. In D. K. Detterman & R. J. Sternberg (Eds.), How and how much can intelligence be increased. Norwood, N. J.: Ablex.

Table 1
Intercorrelations Among Ability and Reference Battery Measures

	FA	APM	DR	LS	R	L
Figural Analogies						
Progressive Matrices	.45					
Diagramming Relationships	.34	.34				
Letter Sets	.36	.25	.36			
Reading	.32	.34	.41	.18		
Language	.32	.35	.36	.21	.72	
Mathematics	.41	.42	.40	.30	.61	.71
Total Test	.39	.41	.42	.26		

Note: All correlations are significant at the $p < .001$ level.

Table 2
Reliabilities for the Systems Thinking Instrument Subscales

Subscale	Items	Alpha
Graphing	7	.51
Graph Interpretation	7	.72
Variables	4	.92
Graph Translation	7	.68
Causality	3	.92
Loops	5	.86
Math Equations	3	.67
Math Graphing	5	.87
Systems ID's	25	.92
Systems Interpretation	10	.87
Total Test	76	.95

Note. n = 179.

Table 3

Intercorrelations of Systems Thinking Instrument Subscales

	G	GI	V	GT	C	L	ME	MG	SI
Graphing									
Graph Interpretation	.57								
Variables	.39	.44							
Graph Translation	.53	.62	.40						
Causality	.40	.56	.33	.54					
Loops	.43	.64	.33	.59	.61				
Math Equations	.49	.51	.27	.46	.53	.42			
Math Graphing	.40	.44	.37	.53	.48	.42	.51		
Systems ID's	.32	.43	.28	.50	.54	.62	.38	.39	
Systems Interpretation	.34	.51	.32	.56	.55	.61	.42	.51	.72

Note. $n = 179$.
All correlations, $p < .001$.

Table 4

Descriptive Statistics for the Systems Thinking Instrument Scales

Subscale	Maximum Score	M	S.D.
Graphing	13	8.94	2.71
Graph Interpretation	20	10.99	5.10
Variables	12	5.30	3.84
Graph Translation	24	11.63	6.09
Causality	9	4.92	2.61
Loops	22	11.54	6.54
Math Equations	6	3.75	2.04
Math Graphing	8	2.75	3.21
Systems ID's	25	13.63	7.20
Systems ' terpretation	46	14.90	10.76
Total Tesc	76	88.36	37.71

Note. n = 179.

Table 5
Course Breakdowns for the Systems Thinking Instrument Scales

Subscale	GPS		Biology		Chemistry		War & Rev.	
	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Graphing	7.39	3.10	8.44	2.31	10.46	1.80	12.00	1.00
Graph Inter.	7.00	3.98	10.43	4.50	13.89	3.76	19.57	1.34
Variables	4.91	3.82	3.70	2.80	6.98	3.95	10.29	2.75
Graph Trans.	8.09	6.12	11.33	5.10	13.64	5.35	21.86	1.07
Causality	2.76	2.56	5.20	2.13	5.87	2.04	8.86	0.38
Loops	5.98	4.78	12.51	5.46	13.68	6.19	21.29	0.95
Math Eq.	2.04	2.36	4.00	1.52	4.57	1.51	5.86	0.38
Math Graph.	0.63	1.72	2.40	3.05	4.27	3.08	8.00	0.00
Systems ID's	10.39	6.52	13.81	7.09	14.82	6.84	23.57	1.62
Systems Inter.	8.96	7.79	15.27	9.51	16.23	9.84	39.71	6.90
Total Test	58.15	29.81	87.10	29.21	104.41	29.96	171.00	9.80

Note. n_{GPS} = 46. $n_{Biology}$ = 70. $n_{Chemistry}$ = 56. $n_{W\&R}$ = 7.

Table 6
First-Order and Partial Correlations Within Subject Area

	First-Order Correlations			Partialling Out APM	
	Pretest	Posttest	APM	Pretest	Posttest
Chemistry (n = 54)					
Systems	.18	.35**	.12	.18	.35**
Pretest		.33**	-.00		.33**
Posttest			.04		
Biology (n = 67)					
Systems	.56***	.40***	.43***	.62***	.39***
Pretest		.32**	-.00		.32**
Posttest			.11		
GPS (n = 36)					
Systems	.25	.43**	.47**	.24	.32*
Pretest		.60***	.08		.61***
Posttest			.36*		

Note. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 7

Correlations Between Achievement and Systems Performance

	Reading	Language	Mathematics	Total
Graphing	.40	.47	.41	.49
Graph Interpretation	.52	.53	.47	.59
Variables	.39	.45	.38	.46
Graph Translation	.51	.51	.50	.57
Causality	.38	.39	.42	.45
Loops	.34	.36	.33	.39
Math Equations	.23**	.29	.45	.35
Math Graphing	.47	.51	.53	.56
Systems ID's	.26**	.20*	.27	.27
Systems Interpretation	.36	.40	.31	.39
Total Score	.53	.56	.52	.60

Note. Chemistry and biology classes only.
* $p < .05$. ** $p < .01$. All other correlations, $p < .001$.
$n_{reading} = 121$. $n_{language} = 122$. $n_{mathematics} = 123$.
$n_{total} = 119$.

Table 8
<u>Dispersion of Students across Treatments, Year 1 to Year 2</u>

Year 1 Course and Treatment	Year 2 Course and Treatment	
General Physical Science - Systems	Biology - Systems	14
	Biology - Traditional	14
	Biology - Traditional	5
- Traditional	GPS - Traditional	1
	Biology - Systems	16
	Biology - Traditional	17
	Biology - Traditional	6
Biology - Systems	Chemistry - Systems	21
	Chemistry - Traditional	34
	Physics - Systems	1
- Traditional	Chemistry - Systems	20
	Chemistry - Traditional	33
	War & Rev. - Systems	1
Chemistry - Systems	Physics - Systems	24
	War & Rev. - Systems	2
- Traditional	Physics - Systems	24
	War & Rev. - Systems	5

Appendix A

Percent Correct on the Advanced Progressive Matrices

Item	*M*	*S.D.*
1-1	.96	.18
1-5	.95	.22
1-9	.66	.47
2-1	.89	.31
2-5	.86	.34
2-9	.84	.36
2-13	.49	.50
2-17	.72	.45
1-3	.90	.30
1-7	.94	.24
1-11	.46	.50
2-3	.91	.29
2-7	.79	.41
2-11	.81	.39
2-15	.63	.48

Note. *n*=370.

Appendix B

Achievement Test Score by Teacher and Course

(In National Percentiles)

Course	Treatment	n	M	S.D.
Reading Subscale				
Biology	Systems	75	70.75	20.73
Biology	Traditional	38	80.45	16.81
Biology	Traditional	34	76.00	17.34
Chemistry	Systems	54	84.89	12.07
Chemistry	Traditional	62	83.16	13.48
War & Rev.	Systems	7	95.86	4.84
Language Subscale				
Biology	Systems	76	62.00	21.46
Biology	Traditional	38	76.00	17.93
Biology	Traditional	35	69.17	19.07
Chemistry	Systems	54	82.98	14.43
Chemistry	Traditional	62	82.90	15.39
War & Rev.	Systems	7	95.00	5.10
Mathematics Subscale				
Biology	Systems	79	68.04	19.83
Biology	Traditional	38	78.42	17.99
Biology	Traditional	32	74.44	19.00
Chemistry	Systems	53	81.89	13.11
Chemistry	Traditional	62	82.14	14.82
War and Rev.	Systems	7	98.14	1.22

Total Test Score

Biology	Systems	74	66.60	20.62
Biology	Traditional	38	79.76	16.52
Biology	Traditional	31	74.94	18.08
Chemistry	Systems	53	85.11	11.58
Chemistry	Traditional	61	84.85	12.02
War & Rev.	Systems	7	97.57	2.22

Note. The two listings for traditional biology reflect the fact that there were two teachers, each with two classes, in the control treatment.

Appendix C

Percent Correct on the Figural Analogies Test

Item	*M*	*S.D.*
1	.87	.34
3	.60	.49
5	.96	.20
7	.92	.27
9	.7[illegible]	.44
2	.90	.30
4	.06	.24
6	.74	.44
8	.35	.48
10	.92	.27

Note. *n*=370.

Appendix D

Percent Correct on the Diagramming Relationships Test

Item	*M*	*S.D.*
1	.92	.27
3	.29	.45
5	.67	.47
7	.45	.50
9	.76	.42
11	.59	.49
13	.40	.49
15	.34	.47
2	.65	.48
4	.28	.45
6	.64	.48
8	.62	.48
10	.76	.43
12	.32	.47
14	.71	.45

Note. *n*=370.

Appendix E

Percent Correct on the Letter Sets Test

Item	M	S.D.
1	.90	.29
3	.96	.20
5	.86	.34
7	.73	.45
9	.66	.47
11	.56	.50
13	.43	.50
15	.10	.30
2	.94	.25
4	.79	.41
6	.80	.40
8	.66	.47
10	.75	.43
12	.70	.46
14	.50	.50

Note. *n*=370.

Appendix F

Item Statistics for the Systems Thinking Instrument

Item	Maximum Score	GPS (n=46)		Biology (n=70)		Chemistry (n=56)		War & Rev. (n=7)	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.
1	1	.54	.50	.87	.34	.84	.37	1.00	.00
2	2	.94	.95	1.43	.79	1.52	.79	1.86	.38
3	2	1.26	.58	1.49	.58	1.36	.67	2.00	.00
4	4	2.22	1.46	2.10	1.32	3.21	1.06	3.57	.54
4a	1	.44	.50	.63	.49	.79	.41	1.00	.00
5	4	1.35	1.61	2.19	1.64	1.91	1.59	3.14	1.07
6	1	.56	.50	.76	.43	.82	.39	1.00	.00
7	1	1.00	.00	.99	.12	1.00	.00	1.00	.00
8	1	.89	.32	.96	.20	.96	.19	1.00	.00
9	2	1.00	1.01	1.91	.41	1.84	.53	2.00	.00
10	2	.48	.84	.76	.84	1.05	.90	1.86	.38
11	2	.56	.91	1.33	.97	1.68	.74	2.00	.00
12	4	2.13	1.92	2.06	1.79	3.12	1.21	3.86	.38
13	4	1.48	1.76	2.49	1.73	3.14	1.28	4.00	.00
14	3	1.26	.95	1.36	1.18	2.41	.93	3.00	.00
15	3	1.26	1.02	1.50	1.09	2.11	.97	3.00	.00
16	3	1.24	1.10	1.34	.95	2.11	.95	2.57	.54
17	4	1.28	1.38	2.00	1.62	2.93	1.35	4.00	.00
18	3	1.33	1.14	.83	.82	1.86	1.10	2.71	.76
19	3	1.15	1.07	.91	.81	1.77	1.08	2.71	.49
20	3	1.24	1.10	.91	.83	1.61	1.02	2.14	1.22
21	3	1.20	1.15	1.04	.88	1.75	1.07	2.71	.76
22a	2	.11	.38	.50	.83	.93	.93	2.00	.00
22b	1	.09	.35	.34	.51	.57	.60	.86	.38
23a	2	.26	.68	.57	.89	.96	.97	2.00	.58
23b	1	.04	.21	.34	.48	.55	.57	1.14	.38
24	2	.13	.50	.63	.94	1.25	.96	2.00	.00
25	3	.96	.94	1.77	.73	2.05	.75	3.00	.00
26	3	.94	.85	1.64	.90	1.91	.82	3.00	.00
27	3	.87	.91	1.79	.78	1.89	.76	2.86	.38
28a	2	.63	.90	1.11	1.00	1.21	.95	2.00	.00
28b	2	.59	.91	1.03	1.01	1.27	.94	2.00	.00
28c	4	.91	1.05	1.94	1.51	2.57	1.58	4.00	.00
28d	5	.80	1.29	2.51	1.83	2.86	1.66	4.71	.49
29a	3	.59	.93	1.51	1.35	1.73	1.29	2.57	1.13
29b	4	1.17	1.61	1.71	1.75	2.07	1.74	4.00	.00
29c	5	.56	.86	2.81	1.80	3.05	1.64	4.57	.54
30	4	1.70	1.35	2.61	1.30	2.68	1.50	4.00	.00
31	4	1.56	1.34	2.66	1.25	2.62	1.48	4.00	.00
32	4	1.35	1.20	1.91	1.34	2.46	1.51	4.00	.00
33a1	1	.52	.50	.23	.42	.45	.50	1.00	.00

Appendix F, Continued

33a2	1	.46	.50	.53	.50	.66	.48	1.00	.00
33b1	1	.41	.50	.63	.49	.66	.48	1.00	.00
33b2	1	.52	.51	.60	.49	.64	.48	1.00	.00
33c1	1	.74	.44	.60	.49	.79	.41	.71	.49
33c2	1	.30	.46	.49	.50	.73	.45	1.00	.00
33d1	1	.74	.44	.73	.45	.71	.46	.71	.49
33d2	1	.56	.50	.64	.48	.62	.49	.86	.38
33e1	1	.59	.50	.59	.50	.70	.46	1.00	.00
33e2	1	.13	.34	.21	.41	.41	.50	1.00	.00
33f1	1	.37	.49	.53	.50	.66	.48	.86	.38
33f2	1	.35	.48	.30	.46	.52	.50	1.00	.00
33g1	1	.41	.50	.56	.50	.54	.50	1.00	.00
33g2	1	.37	.49	.51	.50	.59	.50	1.00	.00
33h1	1	.44	.50	.46	.50	.43	.50	.86	.38
33h2	1	.52	.50	.57	.50	.61	.49	1.00	.00
34a	1	.56	.50	.43	.50	.34	.48	1.00	.00
34b	1	.59	.50	.83	.38	.80	.40	1.00	.00
34c	1	.54	.50	.43	.53	.36	.48	1.00	.00
34d	1	.28	.46	.79	.41	.68	.47	1.00	.00
34e	1	.30	.46	.74	.44	.71	.46	1.00	.00
34f	1	.24	.43	.47	.50	.38	.49	.86	.38
34g	1	.26	.44	.50	.50	.41	.50	.86	.38
34h	1	.09	.28	.73	.45	.70	.46	1.00	.00
34i	1	.09	.28	.73	.45	.73	.45	.86	.38
35a	3	.63	1.18	1.53	1.35	1.55	1.41	3.00	.00
35b	3	.61	.86	1.06	.87	1.02	.88	3.00	.00
35c	3	1.00	1.39	2.19	1.28	2.21	1.28	3.00	.00
36a	7	2.13	1.98	3.17	2.23	3.45	2.30	5.29	1.60
36b	7	1.52	1.86	2.44	2.10	2.73	2.09	5.00	1.63
37	2	.56	.69	.90	.70	.86	.70	2.00	.00
38a	2	.50	.69	.60	.60	.62	.59	1.71	.76
38b	2	.48	.66	.70	.60	.70	.60	2.00	.00
39a	5	.83	1.00	1.07	1.18	1.30	1.54	4.71	.49
39b	12	.70	1.19	1.61	2.16	1.79	2.23	10.00	4.47
39c	4	.74	1.34	.74	1.27	.91	1.27	2.86	1.34