Chapter V
Conclusions and Suggestions for Further Research and Evaluation

5.1 Concluding Remarks

The present study describes our initial attempt to develop some instruments that could be used to measure some of the processes underlying learning competence and to collect some preliminary evaluative data on the ANISA program. Overall, our results suggest that not only are ANISA students developing learning competency skills and perceiving their environment in a very positive way but they are doing so in a way that is equal to or better than their counterparts in a neighboring school area.

Differences favoring the ANISA students were observed for the simple seriation tasks. On the attention and figure ground perception tests the ANISA children performed slightly better than the control students while the reverse was true for cooperation and inflections. The small magnitude of these last four differences implies that the groups should probably be considered equivalent on these tasks. The control group students also exhibit a tendency to excell on the seriation with extension task. This difference may be of some importance. Finally, the pattern of results on the classification subskills indicates that the ANISA students were slightly superior on some aspects of the subskills while the control students were slightly superior on others. Therefore the two groups should be considered essentially equivalent with regard to classification.

For several reasons however we would like to caution the reader

against drawing any strong conclusions about the ANISA model on the basis of our results. One, it was only the first year of the Project and only selected parts of the model were implemented. Two, the implementation was affected by teachers in their first year with the model. We would expect them to do a better job in the second and third year because of the additional experience. Three, the tests that were used, for the most part, were constructed and administered without the advantage of elaborate validation procedures. In addition the tests were not comprehensive in their coverage of the area defined by each process variable. Four, the testing was conducted by relatively inexperienced administrators, but we are neither aware of the importance of this factor for the results nor the direction, if any, the results were biased. Five, it should be recalled that without the aid of baseline data (i.e., data on students at the beginning of the school year) it was impossible to determine the amount of growth during the year. Six, there were many areas in which it was not practical to collect data on during the year. Hence, because of the rather limited scope of the evaluation during the first year many outcomes were not studied and so it would be premature to draw any strong conclusions until data becomes available.

We expect that the majority of the problems raised above will be corrected during the second year of the Project.

5.2 Evaluation and Research Plans for Year II and Year III

During the first year of operation of the ANISA program, the evaluation staff was primarily concerned with the preliminary development or selection of the measurement instruments necessary for carrying out the first two purposes presented in Chapter I. These instruments were used to collect some preliminary data relevant to these two purposes. During the second and third year of operation, it will be necessary to consider methods of refining and extending the evaluation activities. The evaluation activities that should be considered fall into six categories:

a. Definition and measurement of process goals,

b. Definition and measurement of content goals,

c. Experimental design,

d. Definition and measurement of desired teacher behaviors,

e. Basic research,

f. Parental reaction to and support of the program.

In the remainder of this section we will provide a discussion of each of the six categories.

a. Definition and Measurement of Process Goals

The test development and selection activities resulted in tests of seven of the processes underlying learning competence. As there are far more than seven processes posited in the model, the test development and selection activity probably should be extended. The existing tests should be refined in light of the experience with the tests this year.

Test development and selection are concerned with processes which

may loosely be characterized as being at two different stages of conceptualization. Processes such as classification and seriation have stimulated a large amount of research concerned with their development and as a result the desired outcomes, in a developmental sense, may be fairly well specified. Since the research permits the specification of these aspects of the curriculum in behavioral terms, the results of these tests provide useful information about the development of the child. Processes such as attention and verticality have seemingly not enjoyed large research efforts and therefore, it is more difficult to chart a developmental course and concomitantly to specify outcomes in behavioral terms. Consequently, it is difficult to interpret the results of these tests in terms of a child's development. We conceive of two strategies that are available to confront this problem. The first is to conduct basic research on the development of the process. The fruition of that strategy, however, will probably not be immediate. In the interim, for a process like attention which is categorized as an aspect of volitional competence, it may be useful to consider student progress on other processes underlying volitional competence such as goal setting, self-arousal, perseverance, or effecting closure. Although we may know little of the development of these processes, they may be more amenable to definition in meaningful behavioral terms. If this is so, measurement procedures can be devised and the results of these procedures will be interpretable.

Implicit in the above discussion is the point of view that evaluation of student progress would be served by specification of the expected

outcomes of the curriculum by behavioral objectives. This would permit the matching of tests to objectives to aid the interpretation of test results and also permit on-going evaluation of student progress. Moreover, individualization of the curriculum in a pacing and location sense would be facilitated for those processes that have invariant developmental schemes. Mastery of objectives could be assessed and students could be placed in appropriate portions of the curriculum.

The proposed extension of the process testing program has implications for comparative evaluation if the number of processes becomes large. Our experience this year has demonstrated the large time investment necessary to test these processes. With a larger number of tests, the problem will be acute. The time investment for ANISA students will increase but as importantly, it will probably become more difficult to find a control school system that will allow extended access to their students. One way of reducing this problem is to develop tests that may be administered to a group, but this approach presently does not seem feasible with young children. A second method would be to eliminate the comparative evaluation with regard to the process goals. A third approach would be to explore the possibility of randomly sampling the processes and making inferences to all the processes. A fourth approach would be to systematically sample objectives according to priorities, although this would probably limit the inferences that could be made concerning all objectives. A final approach would be to explore the possibility of item – examinee sampling, which would result in a group mean score for each process. It should be possible to make comparisons

with a control group on the basis of this statistic.

The use of tests during the year also has implications for comparative evaluation. If the same tests are used during the year and in the year end assessment and furthermore, ANISA students perform in a superior manner, the results will be subject to the criticism that the ANISA students learned the correct test responses rather than generalized skills. Therefore, the need for parallel test forms is indicated if a comparative evaluation of process goals is pursued. A comparison of the cost of producing the parallel tests with the benefits derived from a comparative evaluation should be attempted.

b. Definition and Measurement of Content Goals

As described in chapter one, the ANISA content curriculum includes four facets: the physical environment, the human environment, the environment of unknowns and unknowables, and the self. In addition, there are three symbol systems which mediate the assimilation of the content curriculum and the mastery of the process curriculum: Mathematics, language, and the arts. During the first year, evaluation of the content curriculum was not pursued in our study. It will be desirable during the second and third year, to develop a testing program that reflects the ANISA content curriculum goals. One way of effecting this change is to utilize the various commerically available standardized tests. This approach is potentially useful only in those aspects of the curriculum for which standardized tests exist. However, even for those aspects of the content curriculum, there is some reservation expressed in the literature concerning the usefulness of these

instruments. With regard to language arts, Moore and Kennedy (1971)
write, "English instruction at the elementary school level has been in
a state of flux for more than a decade," and that as a result, most
standardized tests are concerned with, "... only a fraction of the
total area of language arts." Furthermore, they criticize standardized
tests for measurement of isolated language skills and for totally eschew-
ing oral composition and literature.

Given this criticism of standardized achievement tests, it will be
worthwhile to consider the possibility of stating the parts of content
curriculum to which the criticism applies in terms of behavioral objec-
tives. Criterion-referenced tests would then be constructed to measure
these objectives. One of the major difficulties in this approach is
the time consumed in simply writing the behavioral objectives. This
difficulty may be alleviated to a great extent by using the various
objective and item banks available, and by reviewing standardized
tests to choose items and/or subtests that reflect the content curriculum
objectives.

As noted in the last section, the time consumed by testing the proc-
esses and the need for parallel forms of tests of processes may preclude
comparison of the ANISA students with a control group. As the content
curriculum seems more amenable to paper and pencil tests, group testing
is feasible. In addition, the method of item-examinee sampling would
further reduce the necessary testing time. It would, therefore, be
possible to make control group comparisons on the curriculum content in
order to provide evidence of the models effectiveness in relation to other

educational models. This should be considered in any decision regarding the usefulness of a comparative evaluation concerned with the content goals.

c. Experimental Design

The experimental design used in this study is known as a static-group comparison design (Campbell & Stanley, 1966) and consists of making observations on the experimental and control groups at the end of the school year. This design fails to control for a number of threats to the internal validity of the experiment. Perhaps most importantly, we simply do not have knowledge of the relation between the experimental and control groups at the beginning of the study and so cannot clearly interpret the results. In order to rectify this situation, we need to implement an experimental design that collects both pretest and posttest data on an experimental and control group.

Two designs seem potentially most useful. These are the Non-Equivalent Control Group Design and the Separate-Sample Pretest-Posttest Control Group Design (Campbell & Stanley, 1966). These designs seem to be equivalent with regard to internal and external validity in this case, but because of time constraints, the introduction of sampling may make the second more useful.

d. Definition and Measurement of Desired Teacher Behaviors

One aspect of the present evaluation, involved developing a measurement procedure for investigating the congruence of the model as implemented with the ANISA principles for designing an environment to foster learning

competence. As noted in Chapter III, this procedure focused on the physical environment and the "ground rules" operational in the learning space. These "ground rules" are primarily concerned with the children's behavior. Presumably, these behaviors will foster learning. It would seem equally important to define desired teacher behaviors and attempt to assess the implementation of this aspect of the model. We acknowledge that this was done to some extent this year via videotaping of the teachers classroom activities. However, the inherent potential for reactivity in this situation and limited time sampling are weakness in the procedure that makes a more systematic approach seem desirable. Such an approach involves defining the relevant teaching behaviors, training observers, and sampling occasions and teaching situations for observation. Some thought also should be given to reducing the reactive nature of the rather transparent purposes of these observations.

e. Basic Research

Especially in the process portion of the curriculum, the ANISA model depends critically on knowledge of developmental sequences of the various processes and knowledge of the relative importance of the various processes. While existing child-related research may provide some answers, it is clear that the research presently cannot provide all the answers. Moreover, the dynamic sociology of research defines certain problems as primary research targets at certain periods of time, and so it is unlikely that research on all the processes relevant to the ANISA model is being conducted at this time.

These remarks imply that it is critical for the ANISA staff to begin

to formulate the theoretical questions most relevant to the model and to begin to conduct systematic research on these questions.

f. Parental Reaction to and Support of the Model

With the assumption that parental awareness of and support of the goals and methods of the ANISA model would facilitate the development of learning competence, it seems worthwhile to investigate parental reaction to the model. While an extensive investigation of parental reaction does not seem feasible, there may be certain aspects of the model for which it is important to solicit support outside the school, and the effort to orient parents to this contingency and evaluate the orientation process may be beneficial.

5.3 A Final Note

In this chapter we have presented six areas in which evaluation activities may be extended or begun in the next two years. Within certain areas, we have sketched out some of the alternative avenues that might be pursued. For all areas, except perhaps the investigation of parental reaction, if any suggestions are to be implemented, work must begin in June, 1974. Moreover, the suggestions all imply that the ANISA staff members become more involved in conceptualizing measurement devices and in making evaluation decisions. These two desideratas should be considered in making decisions about the evaluation activities that will be pursued in the future.