Continuing the search for the "perfect evaluation package" with which to assess the Bay Area Writing Project model, members of the BAWP staff and the evaluation unit spent January contacting Title IV-C writing projects in California. Because each of these sites is required to evaluate its program as a condition of receiving Title IV-C funds, state administrators of IV-C grants have expressed interest in a systematic state-wide evaluation plan. Although it is too late to develop such a plan for 1978-79 projects, the evaluation staff gained several insights which may be useful to future IV-C writing projects, and other sites.

Of 19 sites contacted, 16 had planned full scale evaluations consisting typically of holistic assessments of student writing and attitude questionnaires from teachers and/or students. Many of these sites had questions about holistic assessment, which in some cases was doubling as competency testing as well as program evaluation. (See the article in the next newsletter on "Uses of Holistic Assessment.") The remaining three sites welcomed offers of assistance in planning their evaluations.

As reports came in, several common problems emerged. It also became clear that reliance on holistic assessment and attitude questionnaires was causing projects to neglect several other fruitful sources of information about the effects of their program. Future evaluators of writing project sites might consider the following suggestions for improving evaluation components:

**Improving Current Evaluation Practices**

1. **Holistic Assessment.** Writing samples are a more valid measure of student achievement in writing than any standardized objective test. As training procedures for readers improve, the scoring of samples becomes increasingly reliable. But, like any other test, no matter how well administered, the scores it yields must be interpreted carefully.

Here are some suggestions for strengthening holistic assessment of writing:

A) Establish base-line date. How do students *usually* improve in the course of a year? How did last year's students improve? How much do students improve who are not instructed by BAWP-trained teachers?

B) When comparisons are made between scores of one group and scores of another, allow for differences in the *initial ability* of students in these groups. Don't expect poor writers to change in the same ways good writers may. (Adjustment for individual differences can be made statistically. The "method of Adjusted Averages" is a simple effective approach that does not require computer technology.)*

C) Keep records of what changes teachers introduced into their teaching after the BAWP programs: peer group responses, sentence-combining, etc. Evidence of such changes strengthens the possibility that changes in student performance are due to impact of in-service, rather than other influences.

D) Collect at least TWO writing samples, in different modes, at any one testing. Evidence is that students respond differently to different rhetorical tasks. Certainly the training of teachers addresses itself to many different kinds of writing. Improvement of instruction in exposition or argument, for instance, will remain invisible if only personal essay writing is tested.

In addition, the particular writing stimulus, the time allowed for writing, the opportunity (or lack thereof) for rewriting or pre-drafting, the nature of the topic, the rubric (instructions to readers), all limit interpretation of the scores. How, in fact, does the tester define "writing" when he sets out to measure "writing ability?"

II. **Short-term vs. long-term effects of program.** Title IV-C assessments and other attempts to evaluate a writing project in the space of one year are faced with the problem inherent in measuring short-term effects of long-range programs.

A) It is difficult to determine short-term effects emanating from a specific program, except in the most controlled study.

B) Student performance may not improve dramatically in one year, even with improved courses at each level. There is a limit to how much any given student can improve in a short time. What is more likely to be visible is that students who have gone through a four-year program of improved instruction will, on completion of their second, third, and fourth years, be progressively further ahead than comparable students in the old program.

C) Short term effects may be a result of the testing itself. Researchers are familiar with an effect called "regression to the mean." If a student performs at the extreme end of the scale one time, he is likely to move toward the center of the scale on his second performance. Teachers of very poor writers may discover that their students improve noticeably as a group, but this result may be inflated by regression to the mean. Similarly, teachers of top students should not be mystified or disappointed if some of their best writers who perform at the top of the scale on the pre-test, move down a point or two on the post-test.