ASSessment of the Chicago area writing project

Writing Project directors are in agreement that
the experiences teachers have in a summer institute
profoundly change what happens in their class-
rooms, but finding a way to demonstrate this
phenomenon has not been easy. Perhaps the expe-
rience of the Chicago Area Writing Project will be
useful to other Project directors. We have just
completed an assessment of student writing that
has shown significance for the Project group at a
gratifying .001 level. The results of our study provide
the data for our application for validation, currently
in the hands of the Illinois Title IV, Part C evaluators.
With validation will come funding through the
Illinois State Board of Education to take the Chicago
Area Writing Project to other school districts
throughout the state.

Like other writing projects, we held a five-week
summer institute for selected teachers during which
participants wrote a great deal, shared and discussed
their writing in small groups, reflected on, analyzed,
and mapped the writing process, compared their
experience with research studies and current writ-
ing theories, discussed implications for effective
teaching, and generated strategies for the class-
room. They gave presentations appropriate for
inservice work with other teachers and planned
workshop sessions to present to their colleagues
during the school year. Participants were teachers
of all grade levels from one through twelve in
eighteen schools in six widely different school
districts.

Our goal was to set up a research design that
would validly test writing and yet be simple enough
to be carried out within our budget. We wanted,
however, to do more than just look at how teacher
behavior had changed; we wanted to know if
students were actually writing better because of our
teacher training. To find out, we have students write
a story as a pre test and again as a post test, and we
holistically evaluated their writing samples, using
the primary trait scoring system.

Our evaluation showed that students whose
teachers had been trained in our summer institute
made twice as much progress as those in control
schools. Our testing instrument and scoring system
was that used by the National Assessment of Educa-
tional Progress in each of their three writing evalua-
tions. As far as we know, we are the first to use a
NAEP instrument in a pre and post testing of the
same students.

Our research design is a traditional experimental-
control one, a variation of the Solomon Four-
Group Design. We compared the progress from fall
to spring (or fall to January in the case of one-
semester classes) of students whose teachers partici-
pated in the summer institute (T1) with a control
group of students whose teachers had not (T2). A
random sample of students of the thirty one teacher
participants from six Chicago area school districts
were compared with a random sample of students
of twenty four matching teachers chosen from
control schools. Control school selection was based
on matching the overall characteristics of school
communities using criteria NAEP followed for
community-type classification. Empirical data show-
ing equivalency of schools also took into account
racial-ethnic breakdown and profiles of standard-
(Continued on page 2)
last night a very odd-looking bird appeared in the neighborhood.”

Student writing samples were scored using the primary trait scoring system developed by NAEP. This most sophisticated of the holistic scoring systems is reliable for comparing student performance from year to year. Unlike other holistic scoring, it does more than rank the performance of a particular group. A trait is either absent in a given writing sample, or it is present in some degree. The criteria for determining the degree are explicit, so assessors have more than a general impression about the quality of a given piece of writing. They can then compare overall writing performance over time and between groups because the primary trait scoring system yields more absolute information than other holistic scoring systems do.

The primary trait of the Stork Expressive test is: expressing a story. To succeed, a student must write fiction to account for a situation. Thus, the first thing readers of student papers looked for was narrative structures and amplifying detail to account for the given situation (explanation) and to “entertain with a particular view of the world (expression). The techniques of fiction require control of a consistent point of view.” (NAEP Scoring Rationale, “Stork-Expressive,” #0-102010-3)

Two members of the CAWP staff were trained by the same assessors who train NAEP scorers to administer the tests and to read and score the writing samples. Anne Mickles, the Project Evaluator, who was one of those trained, in turn trained the rest of the scorers, having them score training papers until they internalized the scoring system and became consistent in their assessments of the primary trait, which was to express a story. Each paper was assigned a number from 0 through 4, based on this scoring guide:

0. No response; picture instead of a story; illegible, illiterate; misunderstanding of task or writing on a different topic; or I don’t know
1. No evidence of story-telling
2. Little storytelling beyond a gesture at the basic task. Invention of a situation, but flawed in one of these ways:
   a. Plot outline (beginning, middle, end) but little or no elaboration
   b. Rambling details but no plot
   c. Only a beginning
   d. Several separate stories with no connection among them
3. Clear evidence of plot and elaboration with appropriate details. Greater coherence with amplitude than 2, but flawed in one of these ways:
   a. One part of basic plot thinly or inconsistently detailed
   b. Situation established, and plot developed,
but no clear or appropriate closure
c. Completely elaborated plot, but inconsistent point of view, handling of dialogue or management of narration
4. A complete story, amply and appropriately detailed and fully as well as consistently resolved

The scoring of the tests was blind, meaning that the cover sheets that identified students by name, school, teacher, grade, age, class, and sex were removed and coded, so the readers could make no distinction between the target and control group. The pre and post tests were also mixed together and scored at the same time. All students in each class took the test, but because of the large number of students tested (1,600), only half of the tests were scored, and the selection was random.

The results of all this evaluation were impressive, as this table shows:

Mean Writing Score Breakdown by Treatment Group and Level Comparison of Pre test, Post test and Difference Scores

<table>
<thead>
<tr>
<th>Grades</th>
<th>Pre test</th>
<th>Post test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>T2</td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td>Grades 1, 2</td>
<td>0.76</td>
<td>0.93</td>
<td>1.66</td>
</tr>
<tr>
<td>Grades 3, 4</td>
<td>1.81</td>
<td>1.88</td>
<td>2.30</td>
</tr>
<tr>
<td>Grades 5, 6</td>
<td>2.31</td>
<td>2.25</td>
<td>2.56</td>
</tr>
<tr>
<td>Grades 7, 8</td>
<td>2.54</td>
<td>2.58</td>
<td>2.85</td>
</tr>
<tr>
<td>Grades 9, 10</td>
<td>3.08</td>
<td>3.12</td>
<td>3.22</td>
</tr>
<tr>
<td>Grades 11, 12</td>
<td>2.87</td>
<td>2.94</td>
<td>3.01</td>
</tr>
<tr>
<td>TOTAL X</td>
<td>2.57</td>
<td>2.53</td>
<td>2.85</td>
</tr>
</tbody>
</table>

T1 = Project Group    T2 = Control Group

Though there are no national norms for growth in writing, it is striking that students of Project teachers improved on the average more than twice as much as controls, across grade levels. Since scores were in whole numbers only, the .26 average change for Project students reflects that over one-third of students' scores went up one whole point on the five-point scale, whereas fewer than one-fourth of the control students' scores went up and more control students than Project students' scores went down. Total growth was more in lower than in upper grades, but this was not surprising, considering that elementary children saw Project teachers all day, while upper grade students saw trained teachers only fifty minutes per day, and many high school classes lasted only one semester. Student gains held up consistently in different target schools.

Our research results were not only significant statistically but were related to valid educational goals because they 1) evaluated students' actual writing; 2) evaluated content, and did so in an objective, defined, reproducible way using a nationally verified instrument and the primary trait scoring system developed by NAEP; and 3) reflected in the testing instrument the same broad range of writing competence stressed in teachers' training—generation of ideas, elaboration of detail, organization, and response to audience.

We would hope that similar results could be achieved by a replication of this evaluation in other Project sites. We also hope that our demonstration of effectiveness will provide further evidence for funding sources that Writing Projects do achieve their aim of improving student writing performance.

By Betty Jane Wagner and Steven Zemelman, Co-Directors and Anne Mickles, Evaluator, Chicago Area Writing Project