This study investigated whether training in reflection improves the quality of beginning preservice teachers' pedagogical thinking. The first problematic element for the reflecting student is becoming aware of essential aspects and evaluating action. The study compared the effects of two types of reflective training on two groups of students. One group (N=30) was taught to engage in structured, systematic thinking about reflection; the other group (N=29) engaged in unstructured journal writing as a means of encouraging reflection. The effects of the training were assessed by measuring the quality of pedagogical thinking. The Taxonomy of Teacher Reflective Thinking Rating Scale (TTRT) and the Description, Interpretation, Evaluation, Planning Rating Scale (DIEP) were used as pre- and posttest instruments. Before training, all students were given the TTRT as a pretest. Both groups received six 1-hour training sessions approximately 1 week apart. One group wrote DIEP reflections and the other wrote daily journal reflections. At the end of 14 weeks, the Reflection Attitude was administered, as well as the TTRT, as a delayed posttest. Results did not indicate that the treatment alone caused student gains from the pretest to the posttest. The amount of structure in reflective training did not appear to be a significant cause in changing pedagogical thinking. (JD)
The Effect of Structured Training vs Less Formal Journal Writing on Quality of Thinking, Classroom Teaching Performance, and Attitudes Toward Reflective Teaching During Preservice Teacher Training

Paul F. Cook
Janet R. Young
Beverly R. Cutler

Brigham Young University
Department of Elementary Education
Provo, Utah 84602

The Effect of Structured Training vs Less Formal Journal Writing on Quality of Thinking, Classroom Teaching Performance, and Attitudes Toward Reflective Teaching During Preservice Teacher Training

"Reflection"... is perhaps the most written about single idea in education today (Houston, 1988, p. 7). Dewey (1904) believed "experience plus reflection equals growth." Dewey makes a distinction between the undesirable "routine" action of the nonreflective teacher and the "reflective action" of the professional (Dewey, 1933, p 17).

Pollard and Tann (1987) define a "reflective teacher" as "one who constantly questions his or her own aims and actions, monitors practice and outcomes, and considers the short-term and long-term effects upon each child." Shulman (1987) describes the act of reflection as: "... what a teacher does when he or she looks back at the teaching and learning that has occurred, and reconstructs, reenacts, and/or recaptures the events, the emotions, and the accomplishments. It is that set of processes through which a professional learns from experience." (p 331)

Many teacher training institutions are now including a "reflective teaching" component as part of the preservice teacher education curriculum (Cruickshank, 1987, Korthagen, 1985; Zeichner and Liston, 1987). Research evidence of the effectiveness of reflective teaching methods is preliminary and inconclusive. Nevertheless, most educators believe in reflection (Fellows and Zimpher, 1988). A number of issues related to teaching reflection need experimental investigation. For example, it would be useful for educators to know the result of reflecting narrowly or broadly on the problems of teaching, and whether reflectiveness is a developmental characteristic or a skill that teachers can teach. Also, researchers need to develop measures of reflectiveness. Researchers need to establish causal relationships between teacher reflectiveness and outcomes such as improved student learning. Various training models need greater development and more validation.

In this study we tested reflection as a model. Theists agree on the general process of reflective thinking follows. They have described reflection as a cycle with identifiable stages or phases (Korthagen, 1985; Boud, Keogh, & Walker, 1985). The first step the reflector takes is action. The second step is looking back on the action. The third step is becoming aware of essential aspects and evaluating the action. The fourth step is creating alternative methods of action where the action has fallen short of the desired goals. While all steps require substantial skill, the most problematic for the beginning reflector is step three, becoming aware of essential aspects and evaluating the action. In this step the professional diverges dramatically from the novice. The reflector must have a complex set of thinking skills including seeing patterns, establishing criteria, and judging the worth of performance against criteria. She or he must also relate a teaching experience to theoretical concepts. The reflector may name events, identify feelings about what is happening, and decide what effect these feelings are having on the teaching performance. How can educators best teach this reflective process to beginning preservice teachers? Should educators teach it directly and formally, or less formally?

In this study we compared the effect of two types of reflective training. With one group we taught students to engage in structured, systematic thinking about reflection. With another group we engaged students in unstructured journal writing as a means of encouraging reflection. We assessed the effects of these types of training by measuring the quality of pedagogical thinking done by students. Specifically, we investigated answers to the following research questions:

1. Does training in reflection improve the quality of students' pedagogical thinking?
2. Does the amount of structure in that reflective training affect the quality of pedagogical thinking done by students?
3. Does the amount of structure in reflective training affect teaching performance ratings?
4. Does the amount of structure in reflective training affect student attitudes toward reflection?

Answers to these questions may help educators conduct more effective reflective teaching programs.

Method

Subjects

All students (140) enrolled in Elementary Education 310R, Introduction to Learning and Teaching, at Brigham Young University for winter 1989 semester served as the sample population for this study. Fifty-nine students (57 females and 2 males) enrolled in two intact sections of the course during winter semester acted as the target sample for the study. Students register voluntarily for these sections or the department secretary assigns them randomly.

We formed an experimental and a comparison group in each section using a table of random numbers. We combined the experimental groups from both sections and the comparison groups from both sections for analysis purposes. The experimental group had 16 students from section 1 and 14 students from section 2. The comparison group had 15 students from section 1 and 14 students from section 2.

Instruments

Four instruments were used in the study: (1) the Taxonomy of Teacher Reflective Thinking Rating Scale (TIRT), (2) the Description, Interpretation, Evaluation, Planning Rating Scale (DIEP), (3) the Elementary Education Observation Record, and (4) the Reflection Attitude Survey.

Taxonomy of Teacher Reflective Thinking Rating Scale (TIRT). The TIRT was developed at Eastern Michigan University in 1987-88 (Simmons & Sparks, 1989). The scale is designed to measure seven levels of pedagogical thinking. It was developed as a "framework for categorizing seven levels of more-to-less sophisticated examples of reflective pedagogical thinking from students."

Experimenter showed a 10-12 minute video tape of an elementary teacher using a "direct instruction" teaching strategy. After watching the videotape, students wrote for 20 minutes about the vignette. Three trained raters assessed the quality of pedagogical thinking evident in the student papers using the TIRT. Two raters, blind to whether the paper was a pretest, posttest, or delayed post test, and blind to the ratings of other raters, scored each paper. If a difference of scoring of more than one level occurred between the two raters, a third rater scored the paper. We used the two closest ratings in computation and averaged equidistant ratings. We computed interrater reliability coefficients for the study using Spearman Rho, obtaining .72 for the pretest, .75 on the posttest, and .81 on the delayed posttest.

Description, Interpretation, Evaluation, Planning Rating Scale (DIEP). This scale was developed to rate reflection papers written in the DIEP format. It consists of seven subscales: (1) objectivity of description, (2) detail of description, (3) interpretation, (4) evaluation, (5) plan, (6) theme, and (7) technical control. A high rating indicates the ability to focus on a theme, record details objectively, and ground all analytical and planning parts on observations made in the description.

We showed both groups a 10 minute video tape of an elementary teacher instructing, and asked students to write a reflection paper using the DIEP format, even though the journal group had received no training in the DIEP method. Two raters, blind to the other rater's judgment, assessed each paper. If scores were more than 10 points apart, a third rater scored the paper. We used the two closest ratings in the data analysis, obtaining an interrater reliability coefficient, using Spearman Rho, of .70.

Elementary Education Observation Record. We measured teaching performance using an average of the midterm and final teaching performance rating received by students in the course.
Reflection Attitude Survey. We measured attitudes toward reflection using a survey developed for the study. The survey contained thirty-five Likert items, scaled from 1 to 5, grouped into five subscales (see Table 3 for list of subscales). The estimated alpha index of reliability for the subscales ranged from .62 to .82. The estimated alpha index for the total score was .92.

Design

We chose a repeated measures design with three repeated measures (pretest, posttest and delayed posttest) and two comparison groups (structured and unstructured). We used a comparison group rather than a control group in the design. We thought it unethical to jeopardize professional growth of students in the initial stage of preservice training by depriving them of reflective training. Instead, we compared two levels of one independent variable.

Procedure

After assigning students randomly to the two experimental groups, we gave all students the TTRT as a pretest. Then we trained both groups, giving each group six one-hour training sessions approximately one week apart. One group wrote DIEP reflections and the other group wrote daily journal reflections. Two instructors taught the reflection methods because training for both groups occurred simultaneously. However, each instructor gave training of each type of training. Following the six training sessions we administered the TTRT as a posttest. For the remaining weeks the students continued their written reflections but received no feedback from their instructors about their writing. At the end of fourteen weeks we gave a delayed posttest (TTRT) and the Reflection Attitude Survey.

Training

Structured Group. Instructors taught students in the structured group to reflect using the Description, Interpretation, Evaluation, Plan (DIEP) method of written analysis (Cook & Cutler, 1989). The DIEP is a structured written approach to the steps in the cycle of reflection: looking back on the action, becoming aware of essential aspects and evaluating the action, and creating alternative methods of action where the action has fallen short. This method provides students step-by-step practice in reflective thinking. Students observe a teaching episode, or recall their own teaching experience, paying particular attention to details that relate to an assigned theme.

In the “Description” section the students record details of teacher behavior and learner response along with aspects of the physical setting that relate to the assigned instructional theme. Instructors taught students to record only specific, significant details as objectively as possible. In the “Interpretation” section the students attach meaning to the observations by labeling, categorizing, and exploring relationships among observations. Then the students make hypotheses about “if . . . then” relationships.

In the “Evaluation” section the students make judgments about what they observed. Instructors taught them to base their judgments on their values, experience, and knowledge, supporting their judgments with their description and interpretation. Finally, in the “Plan” section students apply their interpretation and evaluation in planning for future teaching. Students wrote five of these reflection reports— one every two weeks. Course instructors rated these reports using the DIEP rating scale to provide feedback to the students.
Journal Group. Students in the journal group kept a daily journal of their experiences and concerns. Instructors monitored journal writing and made encouraging comments but did not provide feedback on the quality of student reflections.

When students met weekly for training, they discussed issues and concerns they had recorded in their journals. In the first training session the instructor defined reflection and assigned the students to write daily reflections in their journals. The instructors did not assign themes for reflection, critique the reflections, or determine the topics for the discussions during the training sessions. They did encourage participation by all members of the group in the discussion.

Data Analysis

Data analysis was done on the university's mainframe VACs computer using SPSSx software.

Question 1: Does training in reflection, whether structured or unstructured, improve the quality of students' pedagogical thinking?
We tested the following statistical hypothesis: No significant difference exists in pretest, posttest or delayed posttest scores on the TTRT for either group. We tested this hypothesis using a two-way analysis of variance for a repeated measures design (Kirk, 1968).

Question 2. Does the amount of structure in training affect the quality of pedagogical thinking done by students?
We tested the following statistical hypothesis: No significant difference exists between the structured and journal groups' TTRT scores on pretest, posttest and delayed posttest.

We tested this hypothesis using a two-way analysis of variance for a repeated measures design.

Question 3. Does the amount of structure in training affect teaching performance ratings?
We tested the following statistical hypothesis: No significant difference exists between the structured and journal groups' scores on the elementary education observation record.

We tested this hypothesis using a one-way analysis of variance.

Question 4. Does the amount of structure affect student attitudes toward reflection?
We tested the following statistical hypothesis: No significant difference exists between the structured and journal groups' scores on the reflection attitude survey.

We tested this hypothesis using a one-way analysis of variance.

Results

We asked if a difference exists in pretest, posttest or delayed posttest scores on the TTRT for either group. The analysis of variance test indicated a significant difference. We rejected the null hypothesis. Table 1 shows the means and standard deviations for the results.
Table 1
Means and Standard Deviations for Pretest, Posttest and Delayed Posttest TTRT Scores

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Delayed Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured</td>
<td>3.08</td>
<td>3.68</td>
<td>3.78</td>
</tr>
<tr>
<td>M</td>
<td>.90</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>Journal</td>
<td>3.21</td>
<td>3.87</td>
<td>3.54</td>
</tr>
<tr>
<td>M</td>
<td>1.04</td>
<td>.96</td>
<td>.73</td>
</tr>
<tr>
<td>Total</td>
<td>3.14</td>
<td>3.76</td>
<td>3.68</td>
</tr>
<tr>
<td>M</td>
<td>.97</td>
<td>.94</td>
<td>.83</td>
</tr>
</tbody>
</table>

Table 2 shows the results of the analysis of variance test for the TTRT Scores.

Table 2
Analysis of Variance Results for the TTRT Scores

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td>69.00</td>
<td>57</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Groups (A)</td>
<td>.02</td>
<td>1</td>
<td>.02</td>
<td>3.22</td>
</tr>
<tr>
<td>Subjects within groups</td>
<td>68.98</td>
<td>56</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td>77.50</td>
<td>2</td>
<td>6.70</td>
<td></td>
</tr>
<tr>
<td>Time (B)</td>
<td>13.40</td>
<td>2</td>
<td>6.70</td>
<td>10.94*</td>
</tr>
<tr>
<td>Groups x Time</td>
<td>1.62</td>
<td>2</td>
<td>.81</td>
<td>1.32</td>
</tr>
<tr>
<td>B X subjects within groups</td>
<td>62.48</td>
<td>102</td>
<td>.61</td>
<td></td>
</tr>
</tbody>
</table>

Note: df for B x Subjects within groups were reduced by 10 because 10 missing scores in the matrix were estimated.
*p<.05

Using Dunn's test to make planned comparisons between pairs of means we determined that the pretest scores differed significantly from both the posttest and the delayed posttest (t=4.38, p < .01; t=3.61, p < .01). Thus we concluded that both the structured group students and the journal group students grew significantly in their quality of pedagogical thinking over the course.

Second, we asked if a significant difference exists between the structured and journal groups' TTRT scores on pretest, posttest and delayed posttest. The results of the analysis of variance test indicated no significant difference. We did not reject the null hypothesis.

To determine if students in the DIEP group could write better reflections using the DIEP format than students in the journal group, the experimenters asked both groups to write a reflection using the DIEP format at the same time they gave the posttest. We rated these written reflections using the DIEP Rating scale. A one-way analysis of variance showed a significant difference between the mean ratings of the two groups (F=17.50, df =
1, p < .000). We concluded that the structured group who received training in the DIEP wrote much better DIEP reflections than the journal group who had received no training.

Third, we asked whether a difference exists between the structured and journal groups' scores on the elementary education observation record. The results of the analysis of variance test indicated no significant difference \( F = 1.54, \text{df} = 1, p > .05 \). We did not reject the null hypothesis. The group mean for the structured group was 31.88 with a standard deviation of 2.57. The group mean for the journal group was 33.09 with a standard deviation of 3.52.

Fourth, we asked whether a significant difference exists between the structured and journal groups' scores on the reflection attitude survey. The results of the analysis of variance test indicated no significant difference between group scores on the summary score or on any of the subscales. We did not reject the null hypothesis. (See Table 3 for a summary of results.)

**Table 3**  
**Means, Standard Deviations, and F Tests for the Structured Group and the Journal Group for Reflection Attitude Survey Scores**.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Structured</th>
<th>Journal</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness of Process</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>19.79</td>
<td>19.65</td>
<td>.00 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>4.04</td>
<td>4.07</td>
<td></td>
</tr>
<tr>
<td>Usefulness of Activities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>19.03</td>
<td>19.22</td>
<td>.07 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>4.00</td>
<td>2.89</td>
<td></td>
</tr>
<tr>
<td>Frequency of Reflecting</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>9.38</td>
<td>8.26</td>
<td>3.39 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>1.72</td>
<td>2.53</td>
<td></td>
</tr>
<tr>
<td>Likableness of Reflection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>21.76</td>
<td>21.13</td>
<td>.55 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>3.93</td>
<td>3.07</td>
<td></td>
</tr>
<tr>
<td>Feelings About Reflection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>20.38</td>
<td>21.78</td>
<td>1.96 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>5.05</td>
<td>4.32</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>90.35</td>
<td>90.04</td>
<td>.01 (NS)</td>
</tr>
<tr>
<td>SD</td>
<td>16.18</td>
<td>13.40</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

From the results of our study we cannot say that the treatment alone caused the gains from the pretest to the posttest and from the pretest to the delayed posttest. The study occurred over the course of fourteen weeks, and because we embedded the treatment in the course, we must consider other aspects of history in any conclusions about the gains. We
controlled for history effects, but because we used no true control group we cannot conclude with certainty that reflective training caused the gains. Gains may have been because of the course experience as a whole. However, the findings demonstrated to us that preservice students may change the kinds of pedagogical thinking they do in as little as fourteen weeks of training.

The amount of structure in reflective training does not appear to be a significant cause in changing pedagogical thinking. Students who developed skill in writing the DIEP's scored no higher on the TTRT posttests than those who wrote informal journal entries. Perhaps the effects of structured reflection take longer to appear. Zeichner and Liston (1987) reached this conclusion:

Our experience has taught us that much unlearning has to go on before most students will accept the need for a more reflective approach to teaching. The time devoted to this task, within a 15-week semester, may not allow educators enough time to overcome prior experience and commonly-held expectations regarding the purposes of student teaching (p 303).

The finding that no difference exists between the two groups in teacher performance ratings agrees with a finding by Peters (1985). He reviewed the research on the effectiveness of "Reflective Teaching," citing six studies on reflective teaching using Cruickshank's methods. He concluded that compared to microteaching, reflective teaching produced no differences in student teaching grades. Researchers have yet to relate improved reflectiveness to improved teaching practice.

We thought that students in the journal group would demonstrate a more positive attitude toward the reflection process than those in the structured group. Students sometimes complained of the difficulty of writing DIEPs. It appeared distasteful for them to write to a set of specifications. However, students did not demonstrate this on the attitude survey.

Our results imply that teacher educators, un ' otherwise demonstrated, may use either structured or unstructured reflective teaching for preservice teachers with equal confidence. Since each approach has its strengths, a combination of the two approaches seems reasonable. Educators should give students reflection tools, both formal and informal, that they may use throughout their career.

It may be that the formal DIEP reflection is particularly useful in the early stages of teaching. It is by nature other-driven (course instructor), structured, and focused toward fairly narrow concerns. Educators may find it helpful in early field experiences when students need to focus on specific aspects of teaching. Educators may also find it helpful as they coach students in the reflective process. However, for lifelong benefit, students should do self-motivated reflection directed toward the deep and broad dilemmas of education. The more self-directed, free and open requirements of journal reflection may be more conducive to the accomplishment of this goal.

Apparently students' reflective ability grows slowly. It would be beneficial to conduct longitudinal studies of student reflection over time. Also, it would be beneficial to conduct studies of teachers' reflection at various stages of their professional development. Such comparison of reflective pedagogical thinking (during initial field experience, during student teaching, as novice teachers, and as seasoned teachers) would increase understanding of the reflective growth process.
References


Appendix 16

END

U.S. Dept. of Education
Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991