DOCUMENT RESUME

ED 409 353   TM 026 874

AUTHOR   Ross, John A.; And Others
TITLE   Teaching Students How To Evaluate Their Work in Cooperative
                  Learning: Results of a Collaborative Action Research
                  In-Service.
SPONS AGENCY   Ontario Ministry of Education and Training, Toronto.; Social
                  Sciences and Humanities Research Council of Canada, Ottawa
                  (Ontario).
PUB DATE   Mar 97
NOTE   36p.; Paper presented at the Annual Meeting of the American
                  Educational Research Association (Chicago, IL, March 24-28,
                  1997). Funding also received from the Durham Board of
                  Education.
PUB TYPE   Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE   MF01/PC02 Plus Postage.
DESCRIPTORS   *Action Research; Attribution Theory; *Cooperative Learning;
                  Elementary Secondary Education; Evaluation Methods;
                  Evaluation Utilization; Foreign Countries; Goal Orientation;
                  Information Dissemination; *Inservice Teacher Education;
                  Instructional Effectiveness; Self Efficacy; *Self Evaluation
                  (Individuals); Skill Development; Student Attitudes; Student
                  Motivation; Teacher Attitudes

ABSTRACT
                  Three approaches to in-service development for teachers
(action research, skills development, and materials dissemination) were
compared using a multimethod evaluation design with innovation-specific and
general outcome measures for students and teachers. Thirty-three teachers and
their students participated in a project to teach students how to evaluate
their work. There were no treatment differences on a self-reported use of
evaluation procedures, personal teaching efficacy, or in general student
outcomes (goal orientations, attributions for success and failure, and
self-efficacy). There were two small but statistically significant
differences favoring action research: (1) teachers in the action research
condition scored higher on outcome expectancy because they had greater access
to teachers who had successfully used student self-evaluation to increase
student achievement and motivation; and (2) students in the action research
condition scored higher on attitudes toward evaluation because their teachers
had a better understanding of how to share control of evaluation, a core
teacher function. The modest differences were attributable to the short
duration of the treatments and to the neglect of student cognitions about
self-evaluation. (Contains 9 tables and 72 references.) (Author/SLD)

Teaching Students How to Evaluate Their Work in Cooperative Learning Results of a Collaborative Action Research In-service

John A. Ross
Carol Rolheiser
Anne Hogaboam-Gray
OISE/University of Toronto

2

Teaching Students How to Evaluate Their Work in Cooperative Learning
Results of a Collaborative Action Research In-service[1]

John A. Ross[2]
Carol Rolheiser
Anne Hogaboam-Gray
OISE/University of Toronto

Three approaches to in-service (action research, skills development, and materials dissemination) were compared using a multi-method evaluation design with innovation-specific and general outcome measures for students and teachers. Thirty-three teachers and their students participated in a project to teach students how to evaluate their work. There were no treatment differences on self-reported use of evaluation procedures, personal teaching efficacy, or in general student outcomes (goal orientations, attributions for success and failure, and self-efficacy). There were two small but statistically significant differences favoring action research: (1) Teachers in the action research condition scored higher on outcome expectancy because they had greater access to teachers who had successfully used student self-evaluation to increase student achievement and motivation. (2) Students in the action research condition scored higher on attitudes toward evaluation because their teachers had a better understanding of how to share control of evaluation, a core teacher function. The modest differences were attributable to the short duration of the treatments and to the neglect of student cognitions about self-evaluation.

District-level in-service continues to be severely criticized. Matthew Miles (1995) described it as "pedagogically naïve...a demeaning exercise that often leaves its participants more cynical and no more knowledgeable, skilled, or committed than before" (p. vii). Although a great deal of experimentation in new forms of professional development has occurred, few comparative studies have assessed the outcomes of different methods. In this paper we compare three frequently used strategies (dissemination, skills training, and action research) in the context of a specific innovation (teaching students how to evaluate their work in cooperative learning settings). We were interested in the differential effects of the methods on teachers and students. In each case we attended to outcomes directly relevant to the target innovation and to more general indicators of improvement.

1

Taxonomies of professional development approaches can be generated by cross multiplying positions on such questions as whether in-service agendas should be controlled by reformers or determined by teachers, whether the scope should focus on specific innovations or provide teachers with opportunities to reconceptualize teaching and learning, whether programs should build productive working cultures in schools or be limited to particular instructional practices, whether teachers learn more by receiving research-validated information about instructional behaviors or by inquiring into their practices, and so forth. Among the diverse types that can be identified by working through the interactions of these questions, stand three methods that are most frequently used by school districts to stimulate teacher growth.

Method 1: Dissemination

The most frequently employed method of renewing teachers is an information reception model in which instructional practices believed to be worthy are delivered through written documents. In this paper we describe materials dissemination as a single strategy approach, distinct from professional development programs that provide materials as a complement to other experiences. With this method, teachers generally receive materials and are expected to select specific items relevant to their classroom assignments and to use the examples given as a stimulus for their own creative inventions. Evaluations of this approach have been largely negative (e.g., Berman et al., 1977), especially for highly prescriptive packages that constrain teacher autonomy or require new patterns of interaction between teachers and students. Yet this method continues to be the prevailing approach to teacher renewal.

One reason for the survival of materials dissemination for professional development is that it is the method preferred by many teachers. Compendia of best ideas that are distributed without support, follow-up or accountability allow for the greatest degree of teacher control. Teachers can use pieces of the package without rethinking their conceptions of teaching and learning, enacting elements of the resource at their own pace. Such materials provide raw material for teachers functioning as independent artisans (Huberman, 1992) formulating their craft in relative isolation from peers.

Skills Development

In the skills development approach, trainers help teachers upgrade their skills through modeling, sequenced practice, feedback, and other direct instructional techniques. The goal is high fidelity use. The inadequacies of this method are well known (e.g., Fullan, 1982; Tillema & Imants, 1995). Skills development sessions are generally too short, designed by non-teachers without regard for recipients' felt needs, provide little conceptual grounding, address disembodied skills divorced from curricular context, give insufficient follow up support when teachers attempt to use new knowledge, fail to provide self or external monitoring of use, and ignore the conditions in which teachers work. Yet there is evidence, not found in all studies, that the skills development approach can contribute to teacher learning. For example, Wade's (1984) meta-analysis of 91 studies found that skills development in-service had a large impact on

2

teacher practice (ES=.90). Other reviews (e.g., Bennett, 1987; Sparks & Loucks-Horsley, 1990) provide similar grounds for optimism.

Some versions of the skills development approach shift control from trainers to teachers, on the grounds that it will increase teacher commitment to professional learning (Clark, 1992; Thiessen, 1992). For example, in peer coaching teachers become trainers for each other. When they follow an agenda prescribed at the district level, results have been mixed (e.g., Galbo, 1989 and Gooding, Swift, Schell, Swift, & McCroskery, 1990 found no effects). In other programs peer trainers set their own agendas, although this approach has been difficult to implement (Grimmett, 1987). The rationale for including a peer component is that sharing professional experiences contributes to constructive conflict about images of teaching and learning (Ross & Regan, 1993) and that changes in instructional practice require reforms in teacher culture (Fullan, 1993; Hargreaves, 1994).

Collaborative Action Research

Action research is also a teacher-controlled approach to in-service. It typically provides a context for teachers to describe professional experiences, reflect on the meanings of personal practice, exchange interpretations with colleagues, and experiment with new teaching ideas (Fullan & Connelly, 1990; Grimmett & Erickson, 1988; Kemmis, 1987). Although originally intended as an emancipatory movement to enable disadvantaged groups to become more powerful by using research data and tools to bolster their claims (Schensul & Schensul, 1992), action research is more frequently encountered as a vehicle for professional development. By providing opportunities for teachers to recognize discrepancies between their espoused theories and their practices, design interventions to strengthen their instructional strategies and collect systematic data on effects, action research may increase teacher self-awareness and career maturity. For example, teachers may be more likely to act on data which they collect themselves (Cousins & Earl, 1995). It may also contribute to or fundamentally alter the knowledge base about teaching (Lytle & Cochran-Smith, 1990). Little research has been conducted on the contribution of action research to teacher growth (Loucks-Horsley, 1996), although a few studies have demonstrated powerful effects, often when action research is combined with constructivist in-service (e.g., Bell & Gilbert, 1996; Northfield, 1993).

Teachers are more likely to realize the potential of action research if they participate in research partnerships with trained researchers. Partnerships help overcome such obstacles as lack of teacher skill in research methods, a problem affecting even teachers with formal training in research methods (Green & Kvidahl, 1990). Frequent contact with professional researchers through joint research may strengthen the image of the teacher as researcher generating and using findings to improve practice (Huberman, 1995). Teacher involvement in action research is also limited by lack of time to do research, a problem that can be reduced if collaboration with professional researchers brings additional resources to the enterprise. However, the provision of funding carries with it accountability, and demands for visible products may imperil teacher control and elevate the authority of professional researchers (Noffke, 1997).

3

Designing a Fair Comparison

No study has compared the differential effects of these three methods. Separate studies of each type are difficult to compare because they rarely use similar methodologies or share intended outcomes. Dissemination and skills development approaches have mainly been examined quantitatively, with the occasional use of qualitative data to illuminate findings (e.g., Mathieson, 1992). The outcomes of greatest interest have been specific changes in teacher practice, typically defined by an implementation profile (Hall & Hord, 1987) or program template (Scheirer, 1996) and, in the best studies, student improvements. In contrast, the impact of teacher-controlled in-service, including action research, has been largely determined with ethnographic methods (e.g., Anderson, Herr, & Nihlen, 1994).

To provide a fair comparison we adopted a multi-method, multi-outcome approach. We used both quantitative and qualitative methods, balancing surveys completed by all students and teachers with focus group interviews with a subsample of students. The outcomes we selected were specific measures focused on specific changes in teacher practices, usually associated with evaluations of skills development programs, as well as general measures of teacher development (such as teachers' confidence in their professional abilities), usually associated with action research. Our student measures also had a twin focus. We examined innovation-specific effects, as well as impact on more broadly based outcomes such as motivation for learning.

In our comparison of the three methods we focused on a specific innovation linked to more broadly based reform. The innovation we chose was student self-evaluation because teachers, strongly expressed a need for it. Previous research has found that teachers believe that assessment of student performance is a key professional task on which they need to be more proficient (Bennett, Wragg, Carré & Carter, 1992; Gullickson, 1986; Impara, Plake, & Fagar, 1993; Marso & Pigge, 1992). The movement away from psychometric evaluation approaches to authentic assessment methods has accentuated teacher concerns about their evaluation methods, particularly if they experience conflict between their teaching beliefs and the learning theory implicit in the new assessment paradigm (Briscoe, 1991; Lorsbach, Tobin, Briscoe., & LaMaster, 1992). In addition teacher misconceptions about new assessment techniques abound (Oosterhof, 1995; Ruiz-Primo & Shavelson, 1995).

Student evaluation is particularly problematic for teachers using cooperative learning methods. For example, they have to disentangle individual from collective performances because students who coast on the work of others must be identified, parents want reports focused on their child, and administrators are legally obliged to promote individuals not groups. Even exemplary cooperative learning teachers, confident about other dimensions of their teaching, express uncertainty, guilt, and anxiety about their student assessment practices (Ross, Rolheiser & Hogaboam-Gray, 1995). Educational research provides these teachers with little guidance. A few studies (Archer-Kath, Johnson & Johnson, 1995; Conway, Kember, Sivan & Wu, 1993; Huber & Eppler, 1990; Johnson, Johnson & Stanne, 1990; Ross, 1995a) found that specific evaluation procedures, such as structured peer review of group processes, have a positive effect on student achievement. But these studies are largely unknown to teachers and the findings have not been widely implemented.

4

When we interviewed cooperative learning teachers about assessment we found that teachers were experimenting with student self-evaluation and wanted to learn more about it (Ross et al., 1995). We shared their interest because previous studies have found that teaching students self-evaluation techniques has a positive effect on students' achievement (Arter, Spandel, Culham & Pollard, 1994), self-regulation (Henry 1994; Schunk, 1994, 1995), motivation (Hughes, Sullivan, & Mosley, 1985), and use of mastery-oriented help seeking and help giving learning strategies (Ross, 1995a). We also noted that cooperative learning manuals (e.g., Bennett, Rolheiser, & Stevahn, 1991; Ellis & Whalen, 1990; Johnson & Johnson, 1987) encourage teachers to give students opportunities to evaluate their work and provide classroom-ready tools that guide students' reflection on their progress.

Research Questions

Our research was guided by the general question: "Which approach to in-service, materials dissemination, skill development, or action research, will have the greatest positive impact on teachers and students?" The absence of previous studies comparing the three treatment conditions made it difficult to formulate specific hypotheses about the direction of differences. We anticipated that there might be an advantage for the skills development approach in terms of innovation-specific teacher practices and an advantage for the action research approach in terms of broadly conceived measures of professional growth such as self-confidence.

Method

Sample Our goal was to recruit 36 experienced cooperative learning teachers from elementary and secondary schools in a single district in central Ontario (Canada). Only 25 teachers (the actual number was 26 but two teachers worked as a teaching team in one grade 4-6 classroom) volunteered. These teachers were randomly assigned within schools to the action research and skills development conditions. Pre and post data were obtained from 11 teachers in the first condition and 13 (12 classrooms) in the second. A backup strategy was used to recruit teachers for the materials dissemination condition. A request for participants was issued at a secondary school principals meeting and an elementary panel consultant contacted a number of schools that had not sent participants in the first call. This produced 21 teachers who were sent materials, 9 of whom returned pre and post data.

Table 1 summarizes the characteristics of teachers who submitted complete data. There were few differences between treatment groups. Dissemination teachers were more likely to be female (only one male compared to 4 or 5 in the other treatments,) have less experience (10.44 years compared to 11.30 and 11.46 in the other conditions), and were slightly more likely to be in the elementary panel.. All teachers were full time. Very few had masters degrees.

Table 1 About Here

Innovation Specific Teacher Outcome Instruments The measure of innovation-specific practice consisted of 10 Likert items measuring teachers' self-reported *use of assessment methods*

that are fair, transparent, participatory, and collaborative (e.g., "My students help me interpret assessment results."). The survey was completed at the beginning and end of the project. In addition, we asked teachers to compile portfolios containing instruments and strategies developed or selected from a handbook of resources (Rolheiser, 1996). Teacher reflection sheets (for recording teacher observations about the effects of assessment) were distributed because self-monitoring contributes to teacher change (e.g., Anderson & Roit, 1993; Guskey, 1984; Hoover & Caroll, 1987). The reflection sheets and portfolios were completed inconsistently and could not be used to compare the treatments.

General Teacher Outcome Instruments Teachers also completed a pre/post survey measuring teachers' confidence in their professional practice. It consisted of 16 items from Gibson and Dembo (1984), the most frequently used measure of teacher efficacy (Ross, 1995b). Two scores were produced: *Personal teaching efficacy* measured teachers' expectation that they would be able to bring about student learning (e.g., "When I really try, I can get through to even the most difficult students."). General teaching efficacy measured teachers' expectation that teachers (not necessarily themselves) would be able to overcome external influences that impede teachers' success (e.g., "The amount that a student can learn is primarily related to family background."). The latter measure is usually considered to be an *outcome expectancy* indicating whether teachers believe that current methods of teaching are likely to be successful. In previous research (reviewed in Ross, 1995b), personal teaching efficacy and outcome expectancy have predicted adoption of innovative teaching practices and student achievement.

Innovation Specific Student Outcome Instruments Because there was a great range in grade and subject in which teachers experimented with self-evaluation, it was not possible to design an achievement test suitable in all classrooms without subverting teacher control. Prior to the pre-test there was a practice activity in which students evaluated their work following a simple cooperative learning exercise. We introduced the practice activity in case some students had not previously completed a formal self-evaluation in a cooperative learning setting. Students were assigned to four person groups to brainstorm solutions to a simple problem ("why do students get into arguments at school") and reach agreement on the best reason. After the best ideas were collected on the board, students rated their personal performance on the group task by responding to 4 Likert items (e.g., "I listened to my peers in the activity."). They then completed the pretest surveys. On the post-test students responded to the same surveys in terms of self-evaluations they did during the field test.

The innovation-specific student outcome was *attitudes to self-evaluation*. There were 10 Likert items, administered pre and post, measuring the extent to which students believed self-evaluation to be fair, participatory, and helpful (e.g., "My self-evaluation showed how much I had learned.").

Scores on the pretest survey of attitudes were used to select a subsample of students for focus group interviews in the action research and skills development treatments. Within each class the four students with the highest pretest attitudes toward self-evaluation constituted one focus group and the four with the lowest scores formed another. Each group was interviewed for 25-30 minutes about their feelings and beliefs about self-evaluation (e.g., "what did you like/dislike about

6

self-evaluation?" "what would you change about it?"). Ninety-two focus group interviews were conducted.

General Student Outcome Measures The impact of the in-service on broadly based curricular outcomes was measured with several surveys administered pre and post. The constructs operationalized by these instruments were chosen because there was evidence from previous research that the relationship between self-evaluation and achievement is mediated by a mastery goal orientation (Schunk, 1996), internal attributions for success and failure (Marsh & Young, 1996), and higher self-efficacy (Bandura, 1987).

The *goals orientation* survey consisted of 18 items from Meece, Blumenfeld, and Hoyle. (1988) distinguishing four orientations toward learning tasks: mastery (e.g., "The work made me want to find out more about the topic."), ego (e.g., "I wanted others to think I was smart."), work-avoidant (e.g., "I wanted to do as little as possible."), and affiliative (e.g., "I wanted to help others with their work."). *Attributions for success and failure* consisted of 14 items selected from Vispoel and Austin (1995). It produced four scores: internal attributes for success (e.g., "I did well because I tried hard."), external attributes for success (e.g., "I did well because the activity was easy."), internal attributes for failure (e.g., "I did poorly because I have weak skills in this subject."), and external attributes for failure (e.g., "I did poorly because the teacher didn't understand me."). *Student self-efficacy* consisted of 8 items from Cowen et al. (1991) measuring student confidence in their academic ability (e.g., "How sure are you that things will work out well for you when you have to do an activity for the first time?").

Treatment Conditions The *action research treatment* was a partial re-enactment of the experiences of an earlier group of five teacher-researchers (hereafter described as the CLEAR mentors). The CLEAR mentors had conducted inquiries of their own design in which they developed and implemented strategies for teaching self-evaluation (Ross, Rolheiser, & Hogaboam-Gray, 1996). Together with the their academic partners, the CLEAR mentors devised a four stage strategy for teaching students how to evaluate their work: (i) involve students in setting the criteria on which they will be evaluated; (ii) model the criteria; (iii) give feedback on student understanding of the criteria; and (iv) help students use self-evaluation data to set goals). During the action research condition the CLEAR mentors represented the processes and products of their inquiries in a handbook (Rolheiser, 1996), told their stories in narratives, helped the in-service teachers devise their own research projects for teaching student self-evaluation, and acted as coaches while teachers in the action research condition conducted their studies. Teachers in the action research condition were not expected to replicate the experiences of their predecessors but to use the narratives of the CLEAR mentors as examples to be reconstructed in a different curriculum setting.

Teachers in the action research condition met with the CLEAR mentors on three occasions for three hours after school. In session 1 in January they interviewed each other about their current use of self-evaluation, heard an overview of four stages in teaching self-evaluation, participated in three carousel presentations in which CLEAR mentors described their action research, and received the handbook of strategies. They also brainstormed a plan containing the teacher's purpose for focusing on self-evaluation, the specific changes the teacher wanted to make, and indicators of success. Student and teacher pretest instruments were administered immediately after the in-service.

In session 2 in February action research treatment teachers met in small groups with one of the CLEAR mentors to develop action plans. Each teacher was encouraged to focus on as many stages of teaching self-evaluation as they could and to use the handbook in whatever manner they deemed appropriate. Each group identified questions that it wanted advice on. These questions were addressed in a "Consultant Chair" activity in which teachers sought advice from other CLEAR mentors and members of other teams. Teachers returned to small groups to revise their plans. After session 2 teachers returned to their classrooms to implement their plans. Each teacher was given two half days of release time to work on the project, either alone, with another teacher in the school, or with a CLEAR mentor. In addition teachers received brief oral feedback on the results of the student focus group interviews and later received full transcripts.

In session 3 teachers shared their experiences with mentors and peers by constructing personal metaphors of their progress in the project (e.g., a road map) and displayed self-evaluation materials they created. After this late April meeting, the student and teacher post-test instruments were administered.

The *Skills Development Treatment* was an implementation form of professional development in which the strategies for teaching students how to evaluate their work were presented by academics for high fidelity adoption by teachers. Teachers met after school for three hours on three occasions.

In session 1 in January they interviewed each other about their current use of self-evaluation, heard an overview of the project, and (the main event) participated in an activity designed to sensitize them to the value of self-evaluation. Teachers also identified a partner to work with in their own or an adjacent school. After the session the student and teacher pretests were administered.

In session 2 in February teachers participated in four mini-sessions on how to teach self-evaluation. In each mini-session there was a description of one of the four stages in the model, an illustration (usually based on grade 10 writing skills) of a specific strategy for addressing the stage, small group practice in which teachers applied the strategy to another context, and an examination of portions of the handbook that addressed that particular stage. For example, for the first stage of involving students in setting evaluation criteria the strategy was to have students brainstorm suggestions, negotiate their suggestions with those of the teacher, and use student language to describe the agreed-upon criteria. The strategy was illustrated by a teacher (not one of the CLEAR mentors of the action research condition) describing how she developed rubrics to enable grade 10 students to evaluate their short stories. The practice consisted of teachers in small groups acting out how they would involve their students in setting criteria for work habits. The sections of the handbook that were highlighted consisted of "Sharon's story" (a narrative describing how a teacher used T-charts to involve grade 2 students in setting criteria for CL) and specific instruments to assist in setting criteria. In the final activity teachers selected particular instruments they would use in their own classrooms. After session 2 teachers returned to their classrooms to implement their plans. Each teacher was given two half days of release time to work on the project, either alone or with another teacher in the school. Unlike the action research condition, there were no mentors.

After the session teachers received brief oral feedback on the results of the student focus group interviews and later received full transcripts.

In session 3 teachers shared experiences with one another by constructing personal metaphors of their progress in the project and sharing self-evaluation materials they created. After this late April meeting, the student focus groups were re-interviewed. In mid-April teachers administered the post student surveys and completed the post teacher surveys.

Teachers in the *materials dissemination* condition received a copy of the same handbook as the teachers in the other conditions but were given no explicit direction in how to use it. No after school sessions were held, no mentoring occurred, and no release time was given.

<u>Analysis</u> Student and teacher surveys were scanned using Teleform 4.0 and SPSS files were created. Descriptive statistics were developed for all variables in the study. Variables were normalized using log transformations prior to inferential procedures. The first step in the analysis determined the effects of the treatments on teachers' assessment practice and professional confidence. The second step determined the effects of teachers on students' goal orientations, attributions, self-efficacy, and attitudes to evaluation. In each of these two steps a multivariate analysis of variance was used in which pretest score on the outcome variable was a within-subject factor and treatment condition was a between-subject factor.

The student focus group interviews were transcribed and entered in ATLAS/ti (Muhr, 1995) a qualitative software program for developing codes and classifying text. The coding scheme, shown in Table 2, was developed from the data and organized around the interview guide questions. Student utterances were coded for four possible attributes of self-evaluation (enjoyable, fair, participatory, and useful). For each attribute, codes were developed for agreeing/disagreeing that the attribute could be applied to self-evaluation and the reasons for this belief. An additional set of codes was used to code student suggestions for changes, definitions of self-evaluation, misconceptions, and other responses. The transcripts were reviewed by pairs of coders (discrepancies were resolved through discussion) and interpretive notes were written for each class that described the experiences of students for the pre-negative, pre-positive, post-negative, and post-positive groups. A series of comparisons, between pre- and post-responses, between negative and positive groups, and between the action research and skills development treatments, were used to generate themes.

Results

Quantitative Results

Table 3 displays the unadjusted pre- and post-test means, standard deviations, and reliabilities for the instruments. Two of the student scales (work-avoident goal orientation and external attributions for success), each containing only a few items, were deleted from the study because of low internal consistency.

Table 3 About Here

9

Table 4 displays the unadjusted pre- and post-test means for the teacher outcome variables by treatment condition. There were no pre-test difference on assessment practices [$F_{(2,30)}$=.584, p=.564], personal teaching efficacy [$F_{(2,30)}$=.083, p=.920] or outcome expectancy [$F_{(2,30)}$=.732, p=.490]. In all treatments self-reported use of evaluation methods that were fair, transparent, participatory, and useful increased. Personal teaching efficacy was also higher in all conditions, although outcome expectancy declined in two of the three treatments. Table 5 shows the results of the analyses of covariance for each post-test measure, in which pre-test score was a covariate and treatment was the independent variable. There were no pretest-treatment interactions. The picture is one of consistency over time. The three post-test variables were each predicted by pre-test scores. Only outcome expectancy was significantly influenced by the treatment condition. Teachers in the action research condition were more likely to believe that teachers could overcome factors external to the school that impede student success.

Tables 4 and 5 About Here

Table 6 displays the unadjusted pre- and post-test means for the student outcome variables for the three treatments, using class as the unit of analysis. For attitudes to evaluation, the innovation-specific measure, there were no pretest differences among the treatments [$F_{(2,35)}$=.234, p=.792]. On the posttest, students in all treatments were less likely to believe that the evaluation methods used in their classrooms were fair, participatory, and useful. Table 7 summarizes the results of the analyses of covariance for the student outcomes. There was a pretest-treatment interaction. To explore it further, we bifurcated the sample into a high and low evaluation attitude group based on their pretest scores. We used GLM (General Linear Modeling) and Tukey's HSD procedure to examine treatment effects within each pretest group. We found that students who began the project with relatively positive attitudes toward self-evaluation benefited more from the action research condition than from the skills development condition (mean difference=.3074, p=.028). There were no statistically significant treatment differences for students with lower pretest evaluation attitudes. There were also main effects for pretest and for treatment. Pretest scores significantly predicted all post-test measures. The only statistically significant treatment effect, favoring the action research and materials development conditions over the skills development approach, was on student attitudes toward evaluation.

Tables 6 &7 About Here

For the general measures of student improvement there was one statistically significant pre-test difference. Students in the skills development approach had lower self-efficacy than students in the other conditions [$F_{(2,35)}$=3.32, p=.05; other pretest comparisons not shown]. Table 7 shows the stability of these measures. Each post-test measure was significantly predicted by its pretest. In addition the general outcome measures were highly correlated with attitudes to the innovation-specific measure, attitudes to evaluation (r=.14 to r=.53) and to each other. There were no treatment differences on the general student outcomes.

In summary, the quantitative results showed a small advantage for the action research condition on one of the general teacher outcome measures but there were no differences on the

innovation-specific teacher measure. There was also a small advantage for the action research condition on the innovation-specific student measure but there were no differences on the general student outcome measures.

Qualitative Results

Overall Themes In every focus group, including those made up of students with the lowest scores on the pretest attitude measure, we found students who liked self-evaluation and were able to describe how it was useful to them. For some students, the feedback they received from self-evaluation was more meaningful than feedback they received from their teacher because it was more immediate and more frequent. Some believed it was more valid because students had information, especially concerning the effort they expended, that was not available to the teacher. Self-evaluation provided a mechanism for communicating this information. In addition the final grade received was not based solely on ability but included the effort component if self-evaluations were averaged with teacher judgments. This meant that some students got higher grades than they would have otherwise. Most focus groups also had students who believed that self-evaluation would enable them to do better in the future by helping them detect existing weaknesses that they could remedy.

These positive sentiments about the fairness and utility of self-evaluation co-existed with negative feelings and beliefs. Some students were uncomfortable with self-evaluation because they felt they lacked the expertise to mark their work accurately. They said they did not understand the criteria or could not apply them. There was widespread concern about cheating. Some students thought that it was unfair that dishonest students could inflate their self-evaluation in order to get high marks. Others reported being teased by their peers if they gave themselves a high rating. What was missing in these comments was an understanding of the role played by evidence in triangulating self-evaluations with teacher appraisals. Many students had not made the connection between the criteria generation and modeling activities (if these occurred in their classrooms) on the one hand and the use of these criteria to assess student work on the other. Other indications that students did not understand the process came from concerns that self-evaluation did not count toward the student's grade in some classrooms (i.e., it was discarded if it conflicted with the teacher's judgment) or it counted too little. There were also students who did not feel they participated in making decisions about self-evaluation. For example, some felt they were not involved in setting the criteria or in determining the type of self-evaluation form they completed. Many students said that self-evaluation was a waste of time (because it did not count) or repetitive (because it simply confirmed the teacher's judgment) or boring (because the same form was used all the time).

These student concerns indicated that many students were mystified about how self-evaluation was supposed to work. There were other indicators of widespread student misconceptions that continued through the duration of the project. Some students were unable to define self-evaluation or think of an example. The most common definition of self-evaluation was "marking yourself" without reference to the use of criteria/evidence to judge performance, the relationship between work habits and production, or the relationship between self-evaluation and achievement. Although a few students referred to particular criteria in illustrating self-

evaluation (e.g., "how long we took to write it, the length of it, whether it rhymes and all that stuff"), most of the examples involved recording an achievement level, usually by completing a simple scale (1-5) provided by a specific instrument. In addition some students confused self-with peer-evaluation (e.g., self-evaluation is "having other students evaluate like how you work").

These interview data indicate that students entered the project with a mixture of positive and negative views and with misconceptions about how self-evaluation worked. This information was shared with teachers in the action research and skills development treatments immediately after the pretest focus group interviews. But our field notes from the in-service sessions indicate there was little discussion of student beliefs and how teachers might influence them, for example, by highlighting the best student arguments in support of self-evaluation or overtly confronting misconceptions. The assumption made by teachers, CLEAR mentors, and ourselves was that student conceptions of self-evaluation would become more accurate and their beliefs about its worth more positive as they collected and used self-evaluation data. Student misconceptions might also have arisen because teachers reported that they had students evaluate their work much less frequently than they intended and the extensive debriefings that they had planned were truncated by time pressures to cover the curriculum.

Treatment Condition Differences

To compare the treatments we went through several steps. After we had coded the data and sorted the interviews into the categories of Table 2, we created data summaries for each main code category (i.e., enjoyment of self-evaluation, fairness, utility, etc.) for eight groups consisting of 2 conditions (action research and skills development) X 2 focus group types (positive and negative attitudes) X 2 data collections (pre- and post-test). We then created pre/post summary charts, as illustrated in Table 8. This table summarizes student perceptions of the fairness of self-evaluation, for students placed in the negative attitude groups (based on their pretest attitudes to evaluation survey), in the skills development treatment. The first column in Table 8 lists the reasons students gave for saying that self-evaluation was fair or unfair. The numbers in the table represent locations (beginning lines) in the transcripts of the pre (column 2) and post (column 3) interviews. The information in the top panel of the table suggests that these students became less willing to describe self-evaluation as fair--the number of comments labeling their self-evaluation experiences as fair decline. The top panel also shows that students became clearer about reasons for attributing fairness to self-evaluation—the number giving no reason for their beliefs declined-- and they were less likely to focus on negotiating marks with teachers as a source of fairness. The bottom panel of the table indicates that these students reported as much unfairness on the post-test as on the pre-test. There were some changes in the distribution of the sources of unfairness: fewer concerns about cheating and more concern about giving themselves less than they deserve and lack of training in marking. In the next step we compare pre/post changes between positive and negative focus groups. The last step was to compare between the action research and skills development conditions.

Table 8 About Here

Table 9 displays the final summaries. The first column identifies the main code category, the second identifies the type of group (positive or negative), and the third and fourth columns summarize differences between the action research and skills development conditions. These comparisons favored the action research condition in all but two instances.

Table 9 About Here

When asked about their participation in evaluation decisions, students in the action research condition reported becoming more involved in decision making, particularly in setting criteria ("we made a rubric or whatever it's called" r2tap 406) and developing marking schemes. In addition these students were less likely at the end of the project to indicate that their teachers made all the decisions. In the skills development condition, students' comments suggested their decision making role was no greater at the beginning than at the end of the project. In setting criteria, for example, "it's just what the teacher says...the questions are pretty much the same; usually you're just evaluating the same things: the content, grammar, neatness" [r2tbn 478].

As the project progressed, students in the action research condition were increasingly likely to view self-evaluation as fair. The main reason was that it enabled them to communicate how hard they worked ("the teacher doesn't see everything" r2tap 52), particularly if student-teacher conferences were arranged to negotiate discrepancies between teacher and student evaluations. Participating in the development of criteria also increased fairness because students felt they understood these criteria better than if they had not been involved in their creation. Some students responded to the probe about fairness by emphasizing that self-evaluation enabled students to learn more, a response that increased over time. In contrast, students in the skills development condition became less convinced of the fairness of self-evaluation, giving fewer reasons in the post-test than they gave in the pre-test interviews.

Enjoyment of self-evaluation increased in the action research condition, at least among students who began the project with a more positive disposition toward self-evaluation. They liked it because it gave them feedback on their performance and their abilities, gave them credit for their effort, enabled them to set goals, showed that the teacher trusted them, and provided an opportunity for them to have their say. The findings were mixed in that students in these groups were also more likely to say they disliked self-evaluation at the end of the project. The major concerns were cheating ("people who kind of abuse the system and don't deserve what they get" r2tap 513) and boredom ("repeating the same questions in different forms" r2tap 336). Other students disliked self-evaluation because they did not know how to do it or were too hard on themselves. In contrast, in the skills development condition, positive attitude groups, there were no changes between the pre- and post-test in overall liking or disliking self-evaluation. In the negative attitude groups there were no changes from pre- to post-interviews for the action research condition students. In the skills development condition students gave more reasons for disliking self-evaluation, with particular emphasis on the cheating problem.

Students in the positive attitude groups of the action research condition identified a wide range of uses for self-evaluation. The most important of these uses was the information that it provided about areas needing improvement. For example, "you might not realize what you

15   13

might be weak on and then with a self-evaluation it will say you don't have good study habits or something like that so the next project you could improve." (r1tap 2008) Other students suggested that self-evaluation was useful because it made them try harder, helped them see what they were good at, revealed student effort (especially to teachers), raised student confidence, increased motivation, helped them understand the teacher's thinking (especially the criteria used to judge student work), and contributed to goal setting. Support for these uses increased during the project in the action research condition and declined in the skills development condition. In addition support for a counter-productive argument for self-evaluation, inflated grades ("self-evaluation adds about 5% to your final grade" r1tap 317) declined only in the action research condition.

The differences between the treatments were also visible in the comparisons of the negative attitude groups. In the skills development condition, students were more likely to claim that self-evaluation was useless at the end of the project than at the beginning. The most frequently cited reason they gave was that self-evaluation did not count in determining the student's final grade. Many students indicated that it duplicated the teacher's marking or was discounted if there was a discrepancy because the teacher had more expertise in marking than students ("If I'm marking myself I won't necessarily see it. Like I might think it's good but really it's wrong" r2tbn 1057). Others described self-evaluation as a waste of classroom time that could be more productively spent on other things. Several students reported feeling discouraged after self-evaluation ("if I had spent a lot of time on it and got a really bad grade, you're going to wonder if you should put as much effort in the next time" r2tbn 778). In the action research condition support for these arguments declined from pre to post.

There were also treatment differences in response to probes about changing self-evaluation procedures, although Table 9 indicates that the results were mixed. In three of the four comparisons summarized in the table, students in the action research condition were less critical of the self-evaluation methods used in their classrooms than students in the skills development condition.

The qualitative data confirm the pretest-treatment interaction found in the quantitative data. The beneficial effects of the action research condition were strongest among students selected for their extremely positive views on the pretest survey, although improvements were also observed among students selected for their extremely negative pretest attitudes. The qualitative data also suggest that the skills development treatment had a particularly adverse effect on students who began the project with a negative disposition toward evaluation.

Discussion

Our comparison of three in-service strategies (action research, skills development and materials dissemination) produced five findings. First, there were no treatment differences in the innovation-specific teacher outcome. Teachers' self-reported use of student assessment procedures that were fair, transparent, participatory, and collaborative increased over the duration of the project but the changes were small and affected teachers in each of the conditions equally. The most likely explanation for the size of the changes was the duration of the treatment (eight

14

weeks). Previous studies that attempted to implement fundamental changes in the relationship between teachers and students in the classroom, primarily observations of teachers implementing constructivist teaching (e.g., Mosenthal, 1995; Summers & Kruger, 1994), report that more than a year is required for even partial success. The student interview data collected in our project suggest that teachers found it difficult to share control of evaluation decision making, a responsibility at the core of the teacher's authority. Our second data collection may have occurred before they had figured out how to reconstruct their teaching around shared control.

Second, we found that the outcome expectancies of teachers declined in two of the treatments but not in the action research condition. Our explanation is that the action research teachers had greater access to the models provided by the CLEAR mentors. These mentors were classroom teachers who demonstrated how they had successfully integrated shared control on assessment issues within their teaching. They presented their cases in the workshops, responded to questions, and were available as coaches throughout the project (although these coaches were not used as extensively by the action research teachers as we had planned). The stories these mentors told were available to teachers in the other treatments but only through written cases. This lack of access to credible models and teachers' experiences in enacting the ideals of the handbook may have depressed the expectations of teachers in the other treatments that the approach was feasible.

Third, teachers' expectations of their ability to use self-evaluation in their classrooms to promote learning, as measured by the personal teaching efficacy scale, modestly increased in all treatments (about a third of a standard deviation). Teacher expectancies tend to be highly stable in experienced teachers, unless they choose or are forced to make substantial changes in their work (Ross, 1995b). The modest increases that we found suggest that sharing assessment control with students by teaching self-evaluation may contribute to professional renewal. We suspect that if we had tracked these teachers for a longer period we may have found treatment differences favoring the action research condition. Our speculation is based on previous studies associating higher personal teaching efficacy with greater teacher control of curriculum decision making (Berman et al., 1977; Fletcher, 1990; Moore & Esselman, 1992; Raudenbush, Rowen, & Cheong, 1992), an element that was stronger in the action research than in the skills development approach.

Fourth, on the innovation-specific student outcome measure, the results were slightly better for the action research than the other conditions. Student attitudes toward evaluation declined in all three treatments because, we argue, students' expectations were not realized and their concerns were not addressed. The interviews revealed that most students began the project with positive dispositions toward self-evaluation, believing for example, that it could help them learn better. But students also had many concerns about self-evaluation, for example, that it was easy for dishonest students to cheat. These beliefs were founded on little experience with self-evaluation and several misconceptions about it. As they began to experience self-evaluation activities, such as criteria generation and triangulation of self with teacher judgments, they found that sharing control meant sharing the workload. As demands on students increased, some of their fears were realized and some students discovered new concerns that had not occurred to them before. Students' reappraisal of self-evaluation took place with little teacher involvement.

15

There were few attempts to make the benefits of self-evaluation visible to students and attempts to confront misconceptions and negative feelings were rare. In addition, the effects of teaching self-evaluation in one subject may have been diluted by the experiences of students in all the other subjects of their school day, since most of the classes were on rotary timetables.

The failure to address student cognitions about self-evaluation was a problem in all treatments yet students in the action research condition suffered less from it than students in the skills development treatment. Our explanation is that the action research teachers spent more of their in-service time talking about what self-evaluation is and how it could be introduced into individual classrooms. In addition, the action research treatment modeled shared control of evaluation by teachers and students by showing in the workshops how experts and novices could share responsibility for classroom planning. Teachers listened to the advice of the CLEAR mentors but they were constantly reminded of their autonomy. The mentors avoided the problem (observed by Bencze, 1995 and by Bickel & Hattrup, 1995) that teachers who have reconstructed their practice tend to encourage others to adopt their products but not their process of change. The skills development in-service, in contrast, was primarily a top-down model promoting high fidelity implementation, delivered for the most part by outside experts. Teachers were told to share control in the classroom but they did not see it in the in-service.

Fifth, there were no treatment differences on the general student outcome measures: goal orientations, attributions for success and failure, and self-efficacy. Scores on these scales were virtually unchanged throughout the project. Our explanation is that these measures, all correlated with attitudes to assessment, did not change because students had insufficient experience of self-evaluation to make a difference.

Conclusion

Our findings are suggestive of the relative advantage of action research approaches to in-service. But this project enacted a limited version of action research in which there was little formal training in research methods. The training consisted of receiving a model for doing action research (along with five action research cases conducted by teachers like themselves), advice about data collection such as specific indicators for observing success in the classroom, and information about what their students were saying about self-evaluation practices. They also received assistance in planning their action research projects prior to implementation, but little assistance was given when they were working out the details in their classrooms. In contrast, our previous work with the five teachers who became the CLEAR mentors in this project occurred over a two-year period. The future mentors first interviewed cooperative learning teachers about student evaluation (Ross et al, 1995) and then designed action research projects to use the data they collected to improve their own practices (Ross et al., 1996). In all phases, including implementation in the classroom, there was intensive interaction between the future mentors and three academics. We believe that similar results may have been achievable in this study with more time and intensity of interaction

We conclude our study with renewed optimism about the potential of action research as a vehicle for designing local improvement projects controlled by teachers and assisted by

16

outsiders, in this case, academics. The main thing we learned is that action research takes more time than we had allocated. Teachers needed more time to work out how to accommodate an innovation which involves sharing control of a core teacher function with their existing beliefs about teacher and learner roles. Teachers also needed more time for students to understand what self-evaluation is and how it relates to their learning, in addition to learning how to do it. Our previous activities with the five teachers who became the mentors for this project demonstrated that teaching self-evaluation is a powerful tool for improving student achievement and motivation. What we need to do now is figure out how to share the process and product of these successful teaching experiments with other teachers. Our first attempt showed the potential of action research. What we need to do is refine our intervention, putting more emphasis in the next phase on development of teacher skills in conducting action research and addressing student cognitions about self-evaluation.

References

Anderson, G., Herr, K., & Nihlen, A. (1994). Studying your own school: An educator's guide to quantitative research practice. Thousand Oaks, CA: Corwin.

Anderson, V., & Roit, M. (1993). Planning and implementing collaborative strategy instruction for delayed readers in grades 6-10. Elementary School Journal, 94(2), 121-137.

Archer-Kath, J., Johnson, D., & Johnson, R. (1995). Individual versus group feedback in cooperative groups. Journal of Social Psychology, 134(5), 681-694.

Arter, J., Spandel, V., Culham, R., & Pollard, J. (1994, April). The impact of training students to be self-assessors of writing. Paper presented at the American Educational Research Association, New Orleans.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Bell, B., & Gilbert, J. (1996). Teacher development: A model from science education. London: Falmer Press.

Bennett, B. (1987). The effectiveness of staff development training practices: A meta-analysis. Unpublished doctoral dissertation, University of Oregon.

Bennett, B., Rolheiser, C., & Stevahn, L. (1991). Cooperative learning: Where heart meets mind. Toronto: Educational Connections.

Bennett, S., Wragg, E., Carre, C., & Carter, D. (1992). A longitudinal study of primary teachers' perceived competence in, and concerns about, National Curriculum implementation. Research Papers in Education, 7(1), 53-78.

Berman, P., McLaughlin, W., Bass, G., Pauly, E., & Zellman, G. (1977). Federal programs supporting educational change (Vol. VIII): Factors affecting implementation and continuation. Santa Monica, CA: Rand Corporation.

Briscoe, C. (1991). The dynamic interactions among beliefs, role metaphors, and teaching practices: A case study of teacher change. Science Education, 75(2), 185-199.

Clark, C. (1992). Teachers as designers in self-directed professional development. In A. Hargreaves, & M. Fullan (Eds.), Understanding teacher development (pp. 75-84). New York, NY: Teachers College Press.

Conway, R., Kember, D., Sivan, A., & Wu, M. (1993). Peer assessment of an individual's contribution to a group project. Assessment and Evaluation in Higher Education, 18(1), 45-54.

18

Cousins, J. B., & Earl, L. (1995). The case for participatory evaluation: Theory, research, practice. In J. B. Cousins, & L. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 3-18). London: Falmer Press.

Cowen, E., Work, W., Hightower, A., Wyman, P., Parker, G., & Lotyczewski, B. (1991). Toward the development of a measure of perceived self-efficacy in children. Journal of Educational Psychology, 20(2), 169-178.

Ellis, S., & Whalen, S. (1990). Cooperative learning: Getting started. New York, NY: Scholastic Books.

Fletcher, S. (1990). The relation of the school environment to teacher efficacy. Paper presented at the annual meeting of the American Psychological Association, Boston, MA. (ERIC Document Reproduction Service No. ED 329 551)

Fullan, M. (1993). Change forces: Probing the depths of educational reform. London, UK: Falmer.

Fullan, M. (1982). The consultant. In M. Fullan, The meaning of educational change (pp. 180-192). Toronto, ON: OISE Press.

Fullan, M., & Connelly, M. (1990). Teacher education in Ontario: Current practice and options for the future. Toronto, ON: Ministry of Colleges and Universities.

Galbo, K. (1989). Peer coaching: Preliminary findings regarding the realities of follow through. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. Journal of Educational Psychology, 76(4), 569-582.

Gooding, C., Swift, J., Schell, R., Swift, P., & McCroskery, J. (1990). A causal analysis relating previous achievement, attitudes, discourse, and intervention to achievement in biology and chemistry. Journal of Research in Science Teaching, 27(8), 789-802.

Green, K., & Kvidahl, R. (1990, April). Research methods courses and post-bachelor education: Effects on teachers' research use and methods. Paper presented at the annual meeting of American Educational Research Association, Boston.

Grimmett, P. (1987). The role of district supervisors in the implementation of peer coaching. Journal of Curriculum and Supervision, 3(1), 3-28.

Grimmett, P., & Erickson, G. (1988). Reflection in teacher education. New York, NY: Teachers College Press.

Gullickson, A. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23, 347-354.

19

Guskey, T. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. American Educational Research Journal, 21(2), 245-249.

Hall, G., & Hord, S. (1987). Change in schools. New York: State University of New York.

Hargreaves, A. (1994). Changing teachers, changing times: Teachers' work and culture in the postmodern age. Toronto: OISE Press.

Henry, D. (1994). Whole language students with low self- direction: A self-assessment tool. Virginia: University of Virginia. ED 372359.

Hoover, N., & Carroll, R. (1987). Self-assessment of classroom instruction: An effective approach to inservice education. Teaching and Teacher Education, 3(3), 179-191.

Huber, G., & Eppler, R. (1990). Team learning in German classrooms: Processes and outcomes. In S. Sharan (Ed.), Cooperative learning: Theory and research (pp. 151-171). New York: Praeger.

Huberman, M. (1995). The many modes of participatory evaluation. In J. B. Cousins, & L. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning (pp. 103-111). London: Falmer Press.

Huberman, M. (1992). Teacher development and instructional mastery. In A. Hargreaves, & M. Fullan (Eds.), Understanding teacher development (pp. 122-142). New York, NY: Teachers College Press.

Hughes, B., Sullivan, H., & Mosley, M. (1985). External evaluation, task difficulty, and continuing motivation. Journal of Educational Research, 78(4), 210-215.

Impara, J., Plake, B., & Fagar, J. (1993). Teachers' assessment background and attitudes toward testing. Theory into Practice, 32(2), 113-117.

Johnson, D., & Johnson, R. (1987). Constructive conflict in classrooms and schools. New Brighton, MN: Interaction Book Company.

Johnson, D., Johnson, R., & Stanne, M. (1990). Impact of goal and resource interdependence on problem-solving success. Journal of Social Psychology, 129(5), 621-629.

Joyce, B., & Showers, B. (1988). Student achievement through staff development. New York: Longman.

Kemmis, S. (1987). Critical reflection. In M. Wideen, & I. Andrews (Eds.), Staff development for school improvement (pp. 73- 90). East Sussex, UK: Falmer Press.

Lorsbach, A., Tobin, K., Briscoe, C., & LaMaster, S. (1992). An interpretation of assessment methods in middle school science. International Journal of Science Education, 14(3), 305-317.

Loucks-Horsley, S. (1996). The design of templates as tools for formative evaluation. In M. Scheirer (Ed.), A user's guide to program templates: A new tool for evaluating program content. New directions for evaluation #72 (pp. 5-24). San Francisco: Jossey- Bass.

Lytle, S., & Cochran-Smith, M. (1990). Learning from teacher research: A working typology. Teachers College Record, 92(1), 83-103.

Marsh, H., & Young, A. (1996). The distinctiveness of affects in specific school subjects: An application of confirmatory factor analysis with the national educational longitudinal study of 1988. American Educational Research Journal, 33(3), 665-689.

Marso, R., & Pigge, F. (1992). A summary of published research: Classroom teachers' knowledge and skills related to the development and use of teacher-made tests. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mathieson, S. (1992). An evaluation model for inservice teacher education. Evaluation and Program Planning, 15(3), 255-261.

Meece, J., Blumenfeld, P., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. Journal of Educational Psychology, 80(4), 514-523.

Miles, M. (1995). Forward. In T. Guskey, & M. Huberman (Eds.), Professional development in education: New paradigms and practices (pp. vii-ix). New York: Teachers College Press.

Moore, W. P., & Esselman, M. E. (1992, April 20-24). Teacher efficacy, empowerment, and a focussed instructional climate: Does student achievement benefit? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Mosenthal, J. (1995). Change in two teachers' conceptions of math or writing instruction after in-service training. Elementary School Journal, 95(3), 263-277.

Muhr, T. (1995). ATLAS/ti: Computer aided text interpretation and theory building. Release 1.1E. Berlin: Scientific Software Development.

Noffke, S. (1997). Professional, personal, and political dimensions of action research. In M. Apple (Ed.), Review of Research in Education, 22 (pp. 305-343). Washington: American Educational Research Association.

Northfield, J. (1993, April). (1993). Interpreting some different approaches to teacher education. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Atlanta.

Oosterhof, A. (1995, April). An extended observation of assessment procedures used by selected public school teachers. Paper presented at the American Educational Research Association, San Francisco.

Raudenbush, S. W., Rowan, B., & Cheong, Y. (1992). Contextual effects on the self-perceived efficacy of high school teachers. Sociology of Education, 65, 150-167.

Ross, J. (1995b). Strategies for enhancing teachers' beliefs in their effectiveness: Research on a school improvement hypothesis. Teachers College Record, 97(2), 227-251.

Ross, J. (1995a). Effects of feedback on student behavior in cooperative learning groups in a grade 7 math class. Elementary School Journal, 96(2), 125-143.

Ross, J., & Regan, E. (1993). Sharing professional experience: Its impact on professional development. Teaching and Teacher Education, 9(1), 91-106.

Ross, J., Rolheiser, C., & Hogaboam-Gray, A. (1995). "Something I'm just not comfortable doing": Cooperative learning teachers' feelings about student evaluation. Paper presented at the annual meeting of the Canadian Society for the Study of Education, Montreal.

Ross, J., Rolheiser, C., & Hogaboam-Gray, A. (1996). Teaching students how to self-evaluate when learning cooperatively: The effects of collaborative action research on teacher practice. Paper presented at the annual meeting of the American Educational Research Conference, New York.

Ruiz-Primo, M., & Shavelson, R. (1995, April). Rhetoric and reality in science performance assessments: An update. Paper presented at the annual conference of the American Educational Research Association, San Francisco.

Scheirer, M. (1996). A user's guide to program templates: A new tool for evaluating program content. New directions for evaluation #72. San Francisco: Jossey-Bass.

Schensul, J., & Schensul, S. (1992). Collaborative research: Methods of inquiry for social change. In M. Lecompte, W. Millroy, & J. Preissle (Eds.), The handbook of qualitative research in education (pp. 161-200). San Diego, CA: Academic Press.

Schunk, D. (1995, April). Learning goals and self-evaluation: Effects on children's cognitive skill acquisition. Paper presented at the annual conference of the American Educational Research Association, San Francisco.

Schunk, D. (1996). Goal and self-evaluative influences during children's cognitive skill learning. American Educational Research Journal, 33(2), 359-382.

Sparks, D., & Loucks-Horsley, S. (1990). Models of staff development. In W. Houston (Ed.), Handbook of research on teacher education (pp. 234-250). New York: MacMillan.

Summers, M., & Kruger, C. (1993, September). <u>Long term impact of a new approach to teacher education for primary science</u>. Paper presented at the annual meeting of the British Educational Research Association, Liverpool, England.

Thiessen, D. (1992). Classroom-based teacher development. In A. Hargreaves, & M. Fullan (Eds.), <u>Understanding teacher development</u> (pp. 85-109). New York: Teachers College Press.

Tillema, H., & Imants, J. (1995). Training for the professional development of teachers. In T. Guskey, & M. Huberman (Eds.), <u>Professional development in education: New paradigms and practices</u> (pp. 135-150). New York: Teachers College Press.

Vispoel, W., & Austin, J. (1995). Success and failure in junior high school: A critical incident approach to understanding students' attributional beliefs. <u>American Educational Research Journal, 32</u>(2), 377-412.

Wade, R. K. (1984). What makes a difference in inservice teacher education? A meta-analysis of research. <u>Educational Leadership, 42</u>, 48-59.

Table 1: Characteristics of Teacher Sample

Action Research Treatment

Teacher	Subject Specialty	Panel	Number of years teaching	Masters Degree	Gender	Full-time teacher	Subject of treatment class	Grade of treatment class
1	English, Social Studies, Art	Secondary	4	yes	male	yes	English	11
2	English, Social Studies, Art	Secondary	8	no	female	yes	English	9
3	English, Social Studies, Art	Secondary	5	no	female	yes	English	9
4	English, Social Studies, Art	Secondary	31	no	male	yes	Art	9
5	English, Social Studies, Art	Secondary	17	no	male	yes	Art	9
6	English, Social Studies, Art	Secondary	25	no	female	yes	English	9
7	English, Social Studies, Art	Secondary	10	no	female	yes	Social Studies	10
8	English, Social Studies, Art	Elementary	9	no	male	yes	English	7
9	All Subjects	Elementary	1	no	male	yes	All subjects	6
10	All Subjects	Elementary	3	no	female	yes	All subjects	7
11	All Subjects	Elementary	na	no	female	yes	All subjects	8
			M 11.30					

Skills Development Treatment

Teacher	Subject Specialty	Panel	Number of years teaching	Masters Degree	Gender	Full-time teacher	Subject of treatment class	Grade of treatment class
1	English, Social Studies, Art	Secondary	8	yes	male	yes	English	9
2	Math, Science, Other Language	Secondary	14	yes	male	yes	Business, computers, tech	10
3	Math, Science, Other Language	Secondary	5	no	male	yes	Science	9
4	Math, Science, Other Language	Secondary	10	no	female	yes	Science	9
5	Math, Science, Other Language	Secondary	5	no	male	yes	Math	12
6	English, Social Studies, Art	Secondary	7	yes	female	yes	Social Studies	11
7	Math, Science, Other Language	Secondary	28	no	female	yes	Science	9
8	English, Social Studies, Art	Secondary	18	no	female	yes	English	9
9	English, Social Studies, Art	Elementary	15	no	female	yes	All subjects	6
10	English, Social Studies, Art	Elementary	9	no	female	yes	English	8
11	Math, Science, Other Language	Elementary	4	no	female	yes	English	8
12	All Subjects	Elementary	18	no	female	yes	All subjects	5
13	All Subjects	Elementary	8	no	female	yes	All subjects	5
			M 11.46					

24

Materials Dissemination Treatment

Teacher	Subject Specialty	Panel	Number of years teaching	Masters Degree	Gender	Full-time teacher	Subject of treatment class	Grade of treatment class
1	All Subjects	Elementary	27	no	female	yes	All subjects	3
2	English, Social Studies, Art	Secondary	14	yes	female	yes	English	10
3	English, Social Studies, Art	Secondary	6	no	female	yes	English	9
4	Math, Science, Other Language	Secondary	10	no	female	yes	Other languages	8
5	English, Social Studies, Art	Elementary	5	no	female	yes	English	8
6	English, Social Studies, Art	Elementary	14	no	female	yes	Other languages	8
7	All Subjects	Elementary	2	no	male	yes	All subjects	6
8	Math, Science, Other Language	Elementary	8	no	female	yes	Other languages	7
9	English, Social Studies, Art	Secondary	8	no	female	yes	English	12

M 10.44

Table 2 Student Focus Group Codes

D Definitions of Self-Evaluation
    D 0 no real definition
    D e gives only an example of self-evaluation
    D a gives attributes of self-evaluation

E Enjoyment of Self-Evaluation
    E + likes self-evaluation
    E + r gives reason for liking self-evaluation
    E - dislikes self-evaluation
    E - r gives reason for disliking self-evaluation
    E 0   ambivalent about self-evaluation
    E 0 r gives reasons for ambivalence

F Fairness of Self-Evaluation
    F +   self-evaluation is fair
    F + r gives reason for thinking self-evaluation is fair
    F -   self-evaluation is not fair
    F - r gives reason for thinking self-evaluation is not fair
    F 0   ambivalent about fairness
    F 0 r gives reason for ambivalence

P Participation in Self-Evaluation Decision Making
    P +   participates in self-evaluation decisions
    P + w gives way in which student participates in self-evaluation decision making
    P -   does not participate in self-evaluation
    P - r gives reason why student is not involved

U Usefulness of Self-Evaluation
    U + self-evaluation is useful
    U + r gives reason why self-evaluation is useful
    U - self-evaluation is not useful
    U - r gives reason why self-evaluation is not useful
    U 0   ambivalent about usefulness of self-evaluation
    U 0 r gives reasons why ambivalent about self-evaluation

C Changes
    C   self-evaluation should change
    C w gives way that self-evaluation should change
    NC self-evaluation should not change
    NC r gives reason why self-evaluation should not change

O Other
    M misconceptions
    O other

28

Table 3 Student and Teacher Instruments, Unadjusted Means, Standard Deviations, and Reliabilities

	No. of Items	Pretest			Post-test		
		Alpha	Mean	SD	Alpha	Mean	SD
Student Variables (n=608-621 students)							
Goal Orientations:							
Mastery goals	9	.83	3.61	.68	.83	3.65	.63
ego goals	3	.64	3.11	.94	.66	3.15	.90
Avoidance goals	3	-.05	3.74	.65	-.14	3.68	.63
Affiliation goals	3	.48	3.73	.72	.42	3.64	.66
Attributions:							
Internal success	3	.55	3.87	.66	.61	3.80	.66
External success	4	.37	3.51	.63	.47	3.27	.67
Internal failure	3	.64	2.44	.91	.67	2.51	.91
External failure	4	.71	2.39	.89	.66	2.52	.82
Self-evaluation	10	.77	3.53	.64	.82	3.45	.69
Self-efficacy	8	.79	3.17	.71	.79	3.16	.69
Teacher Variables (n=28-31 teachers)							
Assessment Practices	10	.79	4.01	.74	.80	4.42	.60
Personal Teaching Efficacy	9	.76	4.53	.54	.83	4.70	.49
Teaching Outcome Expectancy	6	.83	3.73	.85	.84	3.56	.99

29

Table 4 Unadjusted Means & Standard Deviations for Teacher Variables by Treatment (n=28-31)

Outcomes	Pretest		Post-test	
	Mean	SD	Mean	SD
Assessment practices				
Action Research	3.80	.70	4.38	.57
Skills Development	4.08	.56	4.48	.69
Materials Dissemination	4.14	.99	4.38	.58
Personal teaching efficacy				
Action Research	4.56	.46	4.69	.47
Skills Development	4.52	.68	4.72	.58
Materials Dissemination	4.53	.44	4.67	.40
Materials Dissemination				
Outcome expectancy				
Action Research	4.00	.45	4.00	.67
Skills Development	3.58	1.17	3.29	1.32
Materials Dissemination	3.63	.69	3.43	.62

30

Table 5 Summary of Analyses of Covariance for Teacher Variables (n=28-31)

Outcomes	Pretest Effect	Treatment Effect	Treatment-Pretest Interaction	Model
Assessment practices	$F(1,25)=34.07$***	$F(2,25)=2.76$	$F(2,25)=2.92$	$F(5,30)=6.92$***
Personal teaching efficacy	$F(1,22)=9.09$**	$F(2,22)=.36$	$F(2,22)=.36$	$F(5,27)=2.22$
Outcome expectancy	$F(1,25)=5.17$*	$F(2,25)=3.39$*	$F(2,25)=3.14$	$F(5,30)=12.66$***

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

31

Table 6 Unadjusted Means & Standard Deviations for Student Variables by Treatment (n=32)

Outcomes	Pretest		Post-test	
	Mean	SD	Mean	SD
Evaluation attitudes				
Action Research	3.53	.30	3.50	.28
Skills Development	3.45	.28	3.32	.23
Materials Dissemination	3.59	.32	3.53	.32
Goal Orientations:				
mastery goals				
Action Research	3.52	.39	3.62	.32
Skills Development	3.56	.21	3.63	.24
Materials Dissemination	3.69	.36	3.66	.33
ego goals				
Action Research	3.08	.20	3.07	.28
Skills Development	3.13	.22	3.23	.12
Materials Dissemination	3.11	.28	3.11	.43
affiliation goals				
Action Research	3.74	.25	3.60	.21
Skills Development	3.75	.27	3.68	.26
Materials Dissemination	3.64	.28	3.65	.20
Attributions:				
internal success				
Action Research	3.80	.20	3.75	.20
Skills Development	3.84	.20	3.78	.18
Materials Dissemination	3.90	.25	3.88	.16
internal failure				
Action Research	2.36	.26	2.36	.37
Skills Development	2.54	.31	2.58	.24
Materials Dissemination	2.40	.36	2.52	.51
external failure				
Action Research	2.36	.27	2.44	.23
Skills Development	2.50	.30	2.61	.23
Materials Dissemination	2.32	.39	2.51	.40
Self-efficacy				
Action Research	3.13	.16	3.08	.19
Skills Development	3.07	.23	3.18	.24
Materials Dissemination	3.28	.23	3.21	.24

32

Table 7 Summary of Analysis of Covariance for Student Post-test Variables (n=32)

Outcomes	Pretest Effect	Treatment Effect	Treatment-Pretest Interaction	Model
Evaluation attitudes	$F(1,26)=26.17^{***}$	$F(2,26)=3.66^{*}$	$F(2,26)=3.91^{*}$	$F(5,31)=8.47^{***}$
Goal Orientations:				
mastery goals	$F(1,26)=30.54^{***}$	$F(2,26)=.73$	$F(2,26)=.68$	$F(5,31)=7.72^{***}$
ego goals	$F(1,26)=7.44^{*}$	$F(2,26)=2.62$	$F(2,26)=2.51$	$F(5,31)=2.89^{*}$
affiliation goals	$F(1,26)=23.20^{***}$	$F(2,26)=.71$	$F(2,26)=.67$	$F(5,31)=5.35^{**}$
Attributions:				
internal success	$F(1,26)=27.66^{***}$	$F(2,26)=.08$	$F(2,26)=.08$	$F(5,31)=6.44^{***}$
internal failure	$F(1,26)=22.92^{***}$	$F(2,26)=1.17$	$F(2,26)=1.24$	$F(5,31)=6.22^{***}$
external failure	$F(1,26)=6.76^{*}$	$F(2,26)=2.10$	$F(2,26)=2.14$	$F(5,31)=3.46^{*}$
Self-efficacy	$F(1,26)=6.43^{*}$	$F(2,26)=.62$	$F(2,26)=.68$	$F(5,31)=3.93^{*}$

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

Table 8 Example of Pre/Post Comparison [Perceptions of the Fairness of Self-Evaluation of Negative Attitude Groups in the Skills Development Condition]

Perceptions of Fairness	Pretest Interviews [r1tbn]	Post-test Interviews [r2tbn]
Self-evaluation is fair, no reason given	81, 477, 862, 872, 1351, 1726, 1933, 1935, 2348, 2352	677, 690, 1558, 2529
Fair because you know how hard you worked and the teacher doesn't	995, 1319, 1939	115, 667, 2047
Fair because you can use self-evaluation to negotiate your mark with the teacher	493, 582, 995, 1024, 1961	
Fair because it gives you feedback on your work	2852	937, 1293, 919
Fair because you know how much you put it into it, regardless of your ability		694, 3114
Unfair because of cheating	74, 876, 892, 1004, 1702, 1712, 1781, 1925, 1954, 2146, 2596, 2602	333, 608, 2252, 2260, 2314
Unfair because some students give themselves lower marks than they deserve	1479, 1483	97, 324, 362, 368, 380, 416
Unfair because students are not trained in marking like teachers are	517, 1471	123, 130, 2309, 2322, 2327
Unfair for other reason	1947	2876

34

Table 9 Summary of Treatment Condition Differences in the Student Focus Group Data

Code Category	Group Type	Pre to Post Changes in the Action Research Condition	Pre to Post Changes in the Skills Development Condition
Participation in self-evaluation	positive attitude groups	Less likely to say that students were not involved in making self-evaluation decisions	no change
	negative attitude groups	more likely to say they were involved in setting self-evaluation criteria	no change
		less likely to say the teacher made all the decisions	more likely to say the teacher made all the decisions
Fairness of self-evaluation	positive attitude groups	more likely to say that self-evaluation was fair, particularly because students know how hard they worked better than the teacher	less likely to say that self-evaluation was fair
		more likely to say self-evaluation is fair because it tells students what they need to improve on	no change
	negative attitude groups	more likely to say that self-evaluation was fair	less likely to say that self-evaluation was fair
Enjoyment of self-evaluation	positive attitude groups	more likely to say they liked self-evaluation	no change
		more likely to say they disliked self-evaluation	no change
	negative attitude groups	no change	more likely to be concerned about student cheating
		no change	More likely to dislike self-evaluation

Table 9 continued

Usefulness of self-evaluation	positive attitude groups	more likely to indicate that self-evaluation is useful and identified a greater variety of positive uses	less likely to say that self-evaluation is useful
		less likely to define usefulness in terms of giving yourself a higher grade	no change
	negative attitude groups	No change	Less likely to say self-evaluation is useful, especially in recognising weaknesses and for improving work
		Less likely to say self-evaluation is useless	More likely to say self-evaluation is useless, especially in terms of discouraging motivation
Change in self-evaluation	positive attitude groups	Less likely to say that self-evaluation should be used more frequently	No change
		No change	Less likely to say that self-evaluation is OK as is
	negative attitude groups	Less likely to say that self-evaluation should be eliminated, less frequent, or optional	More likely to say that self-evaluation should be eliminated or less frequent
		Less likely to call for changes in the instruments used in self-evaluation	No change

36

TM026874

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Teaching students how to evaluate their work in cooperative learning: Results of a collaborative action research in-service

Author(s): Ross, John A., Rolheiser, Carol, & Hogaboam-Grey, Anne

Corporate Source:

Publication Date: March, 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all **Level 1** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

_____ Sample _____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

[✓] **Check here**
For Level 1 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) and paper copy.

The sample sticker shown below will be affixed to all **Level 2** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

_____ Sample _____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] **Check here**
For Level 2 Release:
Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature:

Organization/Address:
OISE Trent Valley Centre
University of Toronto
Box 719, 150 O'Carroll Ave
Peterborough, Ontario K9J 7A3
Canada

Printed Name/Position/Title:
John A. Ross, Professor

Telephone: 705-742-8827

FAX: 705-742-5104

E-Mail Address:
jross@oise.utoronto.ca

Date: March 24/97

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it i publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria ar significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:

Address:

Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and addres

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document be contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2d Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

(Rev. 6/96)