Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998a). Skills training versus action research in-service: Impact on student attitudes to self-evaluation. *Teaching and Teacher Education*, 14(5), 463-477.

Skills Training Versus Action Research In-service
Impact on Student Attitudes to Self-evaluation

John A. Ross*

Carol Rolheiser

Anne Hogaboam-Gray

Ontario Institute for Studies in Education

Of the University of Toronto

*corresponding author

OISE/UT Trent Valley Centre,
Box 719, 150 O'Carroll Ave.,
Peterborough, Ontario
K9J 7A1 CANADA
jross@oise.utoronto.ca

Abstract

Teachers in 23 classes in one district were randomly assigned to two frequently used professional development methods. In the *skills training* treatment expert presenters provided strategies for teaching students how to evaluate their work. Teachers received three training workshops, a handbook of resources (containing case studies, strategies, and student self-evaluation instruments), two half-days of in-school release time, and information about the beliefs of their students about self-evaluation. Teachers were expected to implement the strategies presented. In the *action research* treatment teachers received the same resources but the in-service sessions emphasized the process used by a previous group of (CLEAR) teachers to improve their use of student self-evaluation. The CLEAR teachers delivered most of the in-service and acted as mentors to the teachers in the action research condition. Pre- and post-test student surveys and interviews indicated that the action research condition made a more positive contribution to student attitudes toward self-evaluation than the skills training condition. The advantage of the action research condition was attributed to (1) sharing control in the in-service provided a better model of sharing control in the classroom, and (2) the handbook examples provided to teachers in both conditions were generated in earlier action research projects, making the information more accessible to teachers in the action research condition. The modest student impact of the treatments was attributable to the neglect of student cognitions about self-evaluation in the classroom, the short duration of the project (8 weeks) and the dilution of effects through rotary timetables.

Skills Training versus Action Research In-service

Impact on Student Attitudes to 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). Experimentation with new forms of professional development has occurred but few comparative studies have assessed the outcomes of different methods. In this paper we compare the student effects of two frequently used approaches, skills training and action research, in the context of a specific innovation (teaching students how to evaluate their work in cooperative learning settings).

Theoretical Framework

Skills Training

In the skills training approach trainers help teachers upgrade their skills through the study of theory, modeling, sequenced practice, feedback, and other direct instructional techniques. Coaching is a less frequent addition to the model. The goal is high fidelity use (Joyce & Showers, 1988). The inadequacies of this method are well known (e.g., Fullan, 1982; Tillema & Imants, 1995). Skills training 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 training approach can be effective for transfer of training (Bennett, 1987) and can contribute to student learning. For example, Wade’s (1984) meta-analysis of 91 studies found that skills training in-service had a positive impact on students ($ES=.37$).

Some versions of the skills training 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). The rationale for including a peer interaction 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 teachers' workplace 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; Grimmer & 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 used 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. Most action research reports focus on teacher outcomes exclusively, demonstrating growth in individual teachers (e.g., Bell & Gilbert, 1996) and in knowledge about teaching (Lytle & Cochran-Smith, 1990). Student data are rare. For example, Calhoun and Allan

(1994) found that only 11 of 54 schools engaged in action research provided any evidence of student impact.

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 a researcher who uses findings to improve practice (Huberman, 1995). This image can dilute privacy norms that inhibit teachers from publicly reflecting on their work (Cochran-Smith & Lytle, 1996). When professional researchers share decision making with teachers it communicates the respect of relevant adults that enhances teachers' expectations about their ability to complete the research (Ross, Rolheiser & Hogaboam-Gray, in press, b). Collaboration with professional researchers also can provide resources to give teachers time to do research. 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).

Designing a Fair Comparison

No study has compared the differential effects of skills training and action research methods. Separate studies of each type are difficult to compare because they rarely use similar methodologies or share intended outcomes. Skills training has 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 action research, has been

largely determined with ethnographic methods (e.g., Anderson, Herr, & Nihlen, 1994) focused on effects on teacher knowledge.

In our study we focused on students. To provide a fair comparison we adopted a multi-method approach, balancing surveys completed by all students with focus group interviews with a subsample of students. The study focused on student self-evaluation because teachers in our site expressed a need for PD on this topic. Other researchers have also found that teachers believe they need more training on assessment methods (Bennett, Wragg, Carré & Carter, 1992; Gullickson, 1986; Impara, Plake, & Fagar, 1993; Marso & Pigge, 1992). The movement away from psychometric evaluation approaches to alternate assessment methods has accentuated teacher concerns about how they evaluate (Briscoe, 1991; Lorschach, Tobin, Briscoe., & LaMaster, 1992). In addition, teacher misconceptions about newer assessment methods 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 assessment practices (Ross, Rolheiser & Hogaboam-Gray, in press, a). Educational research provides these teachers with little guidance. A few studies (Ross, 1995a; Archer-Kath, Johnson & Johnson, 1995; Conway, Kember, Sivan & Wu, 1993; Huber & Eppler, 1990; Johnson, Johnson & Stanne, 1990) found that specific evaluation procedures, such as structured peer review of group processes, have a positive effect

on student outcomes. But these studies are largely unknown to teachers and the findings have not been widely implemented.

When we interviewed cooperative learning teachers about assessment we found they were experimenting with student self-evaluation and wanted to learn more about it (Ross et al., in press, a). 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) encourage teachers to give students opportunities to evaluate their work and provide tools to guide students' reflection on their progress. We were concerned that student attitudes toward evaluation deteriorate with school experience: the higher the grade, the more negative the attitude (Paris, Lawton, Turner, & Roth, 1991). Greater emphasis on self-evaluation might enhance student attitudes to evaluation because students view evaluation more positively when they collaborate with teachers in the process (Mabry, 1992).

Research Questions

Our research question was: "Which approach to in-service, skills training or action research, will have the greatest positive impact on students' attitudes to evaluation?" We anticipated that the action research approach would make a greater contribution for two reasons. First, in the action research treatment, control of the in-service was shared with teachers to a greater degree than in the skills training condition. The action research approach provided teachers with a model for sharing control in the in-service which is analogous to sharing control with students in the classroom. Second, the specific methods for teaching student self-evaluation

that were used as examples (in the action research condition) or prescriptions (in skills training) were generated in previous action research projects (Ross et al., in press, b). We anticipated that the knowledge embedded in these methods would be more accessible to teachers working in an action research approach.

Method

Sample Teachers in 23 cooperative learning classrooms (the actual number of teachers was 24 because two worked as a team) volunteered to participate in an in-service on teaching students how to evaluate their work. Teachers agreed in advance to be randomly assigned within schools to either the action research or skills training conditions. Table 1 summarizes teacher characteristics by experimental condition.

Table 1 About Here

Surveys Prior to the pre-test survey students evaluated their work following a simple cooperative learning exercise. We introduced the practice activity in case some students had not previously completed a 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 survey (described below) in terms of self-evaluations they did throughout the duration of the project.

The evaluation attitudes survey consisted of 10 Likert items measuring students’ beliefs about the usefulness of self-evaluation (e.g., “my self-evaluation showed how much I had learned.”). The items were adapted from Paris, Turner, and Lawton (1990) and Wiggins (1993). An earlier study

(Ross, 1995, c) found that a 16-item version of the instrument had adequate reliability ($\alpha=.89$), a single factor solution provided the best fit of the data, survey responses were congruent with interview responses, and the items correlated negatively with age, as predicted by previous research indicating that student attitudes to evaluation decline with school experience (Paris, Lawton, Turner, & Roth, 1989).

Interviews Scores on the pretest surveys were used to select a subsample of students for focus group interviews. In each class the four students with the most positive 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 self-evaluation?” “what would you change about it?”). Ninety-two focus group interviews were conducted.

Treatment Conditions The two treatments were the same in their provision of learning resources, amount of in-service time, in-school release time, field test duration, and intended student outcomes. The treatments differed (relatively rather than absolutely) in expectations for participants (replication in skills training versus construction in action research), in control of the in-service (academics in skills training versus teachers in action research) and in-service presenters (outside experts in skills training versus peers in action research).

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 et al., in press, b). 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 teachers devise their own research projects for teaching student self-evaluation, and acted as coaches while teachers planned and 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.

Action research treatment teachers 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. Each teacher brainstormed a plan containing his or her purpose for focusing on self-evaluation, the specific changes he or she 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 other action research treatment teachers. 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, post-test surveys and student focus group interviews were administered.

The *skills training* treatment was an implementation form of PD in which 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 teachers interviewed each other about their current use of self-evaluation, heard an overview of the project, and (the main event) participated in an experiential 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 surveys and interviews 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, a demonstration (usually based on narrative writing skills) of a specific strategy for addressing the stage, small group practice in which teachers applied the strategy to another context, and a review of portions of the handbook (Rolheiser, 1996) that addressed that particular stage. For example, for the first stage of involving students in setting evaluation criteria, the recommended strategy consisted of having students brainstorm possible criteria, negotiate their suggestions with those of the teacher, and use student language to describe the agreed-upon criteria. A presenter (not one of the CLEAR mentors of the action research condition) demonstrated how a teacher could involve grade 10 students (a role

played by teachers attending the session) in the development of a rubric for evaluating short stories. For practice, teachers in small groups planned 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 and set of storyboards describing how a teacher used T-charts to involve grade 2 students in setting criteria for the social skill “maintaining self-control”) 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.

Analysis of Focus Group Data

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 students agreeing or disagreeing that the attribute could be applied to self-evaluation and the reasons for their belief. An additional set of codes was used to code student suggestions for changes, definitions of self-evaluation, misconceptions, and other

responses. Pairs of coders coded the verbatim transcripts and resolved discrepancies through discussion. 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.

Results

Survey Results

Table 3 displays the unadjusted pre- and post-test means and standard deviations for the attitude survey. The internal consistency of the 10-item scale was adequate (Cronbach's $\alpha=.77$ on the pretest and $.82$ on the post-test). Student attitudes toward evaluation declined slightly in the action research condition and to a slightly greater extent in skills training. Scores were normalized using log transformations prior to analysis of covariance. There was a pretest-treatment interaction [$F(1,19)=5.56, p=.029$]. Inspection of the means (we bifurcated the sample into positive and negative attitude groups on the basis of their pretest scores) showed that there were no treatment differences for students who began the project with relatively negative attitudes toward self-evaluation. In contrast students who were relatively positive initially became less so in the skills treatment condition but continued to be positive in the action research treatment. There was a main effect for pretest [$F(1,19)=11.06, p=.004$]: Classes that entered the project with more positive attitudes to self-evaluation were more likely to have high scores on the post-test than classes that displayed less positive attitudes on the first administration. There was also a main effect for treatment [$F(1,19)=5.19, p=.029$] favoring the action research condition. The effect size, using Glass's formula, was $.36$, indicating that the attitude to self-evaluation of the average student in the action research condition on the second interview was as positive as that of the student at the 64th percentile of the skills training sample.

Table 3 About Here

Interview Results

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 training) 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 4. This table summarizes student perceptions of the fairness of self-evaluation, for skills training students in negative attitude focus groups. The first column in Table 4 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 skills training students became less willing to describe self-evaluation as fair as the project progressed--the number of comments labeling their self-evaluation experiences as fair declined from pre to post interviews. 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 deserved 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 the pre/post trends between the action research and skills training conditions.

Table 4 About Here

Positive Arguments About Self-Evaluation In every focus group—positive and negative, pre- and post-interviews—we found students who liked self-evaluation and were able to describe how it was useful to them. Five arguments occurred most frequently. First, students said that with self-evaluation they had a better understanding of what they were supposed to do because they were involved in setting the criteria. For example,

It's easier for you to get like better because you know what you have to do. Sometimes you get confused about what the teacher says you get marked on. But if you tell her what you want to get marked on, then you understand because you're the one who said you wanted to do that, and then you can get better because you know what you're doing.

Second, students argued that self-evaluation was fairer because it enabled them to include important performance dimensions, such as effort, that would not usually be included in their grade. For example,

[name deleted] is like a very smart kid in our class, and he reads strong or superb. But someone like me might read adequate or maybe about as bad as poor. Like I'm trying as hard as he is, but it's just that we're at different spots.

The third argument was closely related to the second. Students thought that self-evaluation was desirable because it enabled them to communicate information about their performance that the teacher had no way of knowing. For example,

The teacher only knows so much of how much effort you put into it. She has to look over the whole class. You know personally how hard you worked on it and how much time you worked at home or if you were just goofing off.

Fourth, students believed that self-evaluation was helpful because it gave them feedback they could use to improve the quality of their work. For example,

You might not realize what you 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.

The fifth argument was that the feedback students received from self-evaluation was more complete and meaningful than the information they received from the teacher. For example,

I like it because you can like see your good points and your bad points...like when you get a mark back from your teacher you look at what you did right, you don't look at what you did wrong.

There were other arguments offered in support of self-evaluation that were not as widely shared. Some students liked self-evaluation because they could help decide what the criteria for judging student work would be. Some thought that with self-evaluation they got higher marks because their grades were averaged with the teacher's grades ("self-evaluation adds about 5% to your final grade"). In some classrooms in which students experienced peer evaluation, the privacy of self-evaluation was appreciated (only the teacher knew the student's grade). Some students argued that self-evaluation told them things about themselves they would not otherwise know and that self-evaluation could be applied to other areas of life (like sports) in which they wanted to do better. Other students reported that self-evaluation increased their self-confidence because it proved that students can be trusted to judge their work fairly.

Table 5 summarizes the focus group differences between treatment conditions. The largest differences in positive arguments for self-evaluation were about fairness. At the end of the project, students in the action research condition were more likely than they were at the beginning to say

that self-evaluation was fair. The change was visible in both positive and negative groups. In contrast students in the skills training condition became less likely to say that self-evaluation was fair. Action research students, in the positive groups only, were more likely to indicate that they liked doing self-evaluation and to argue that it was useful. Neither the positive nor the negative skills training groups experienced a comparable shift. Students in the action research condition reported becoming more involved in evaluation decision making, particularly in setting criteria. In the skills training condition, students' comments suggested their decision-making role was no greater at the beginning than at the end of the project.

Table 5 About Here

Negative Feelings and Beliefs about Self-Evaluation All focus groups expressed concerns about self-evaluation. Some students argued that self-evaluation was unfair because the procedures allowed some students to cheat. For example, "People could take advantage of it and just mark all perfect when it's really not to their best ability." Some students opposed self-evaluation because they lacked the expertise to judge their work. For example,

Well, we really don't know how to mark ourselves. Since it's from our point of view, we may think its good...But with the teacher marking, it's fair, they're trained to mark and they can find what's wrong.

Other students were teased by their peers if they gave themselves a high evaluation. For example,

Well, when you are marking yourself, you don't want to give yourself too high of a mark, because then your friends may think that you're trying to get your mark up.

There were concerns that the ratings students gave themselves did not count enough or were discarded if the teacher had a different grade in mind. For example, "...usually it doesn't count, so there is no point in doing it." Some students rejected self-evaluation because they thought it was a

waste of time, boring and/or repetitive. For example, “Well, it’s boring when you have to fill out the same [self-evaluation form] a whole bunch of times.”

Less commonly, some students believed they were not involved in setting criteria and were not allowed to participate in other decisions about self-evaluation (such as the format of evaluation instruments). Other students did not understand how to use the criteria, especially if they were presented in rubric form (i.e., as a set of performance dimensions with a hierarchy of performance standards for each).

Students in the skills training condition, particularly those in the negative focus group, became more negative about self-evaluation as the project continued. This was not the case in the action research condition. On the exit interviews skills training students were more likely than they had been in the first round of interviews to say that the teacher made all the evaluation decisions. They were also more likely to be concerned about cheating, to dislike self-evaluation, and to argue that it was useless. 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”).

Misconceptions In all focus groups misconceptions about self-evaluation were visible. Some students were unable to define self-evaluation or think of an example. The most common definition was “marking yourself” (i.e., assigning a grade without reference to the use of criteria, evidence to support the judgment or the purpose for doing so). Some students confused self- with peer-evaluation. Other students were uncertain about how discrepancies between self-evaluations and evaluations by the teacher could be resolved. There were no treatment condition differences on these issues.

The qualitative data confirmed 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 suggested that the skills training treatment had a particularly adverse effect on students who began the project with a negative disposition toward evaluation.

Discussion

The survey results indicated that student attitudes toward self-evaluation declined among students who began the skills training treatment with relatively positively attitudes. 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 had 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 they found that sharing control meant sharing the workload. As demands on students increased, many 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. There were few attempts to make the benefits of self-evaluation visible to students and attempts to confront misconceptions and negative feelings were rare.

Even in the action research condition the surveys showed no overall improvement in evaluation attitudes, possibly because of the duration of the project (8 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), reported 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. In addition, the effects of doing self-evaluation in one subject may have been diluted by the experiences of students in the other subjects of their school day, since most of the classes were on rotary timetables.

The student attitude effect size between treatments was as large as the mean student effect size of in-service treatments reported in Wade's (1984) meta-analysis. The qualitative findings, more than the survey results, suggested that positive changes in student attitudes to evaluation were beginning to occur in the action research condition to a greater extent than in the skills training treatment. Our explanation is that 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 teachers were constantly reminded of their autonomy. The mentors avoided the problem (observed by Bencze, 1995 and by Bickel & Hatrup, 1995) that those who have reconstructed their practice tend to encourage others to adopt the products but not the process of their change. The skills training in-service, in contrast, was primarily a top-down model promoting high fidelity implementation, delivered for the most part by outside experts. Teachers were encouraged to share control in the classroom but they did not see it in the in-service.

Conclusion

This study is unique. None other has compared the impact of these two frequently used in-service approaches, skills development and action research. 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 limited information about what their students were saying about self-evaluation practices. (The complete focus group interviews transcripts were not available until after the project.) They also received assistance in planning their action research projects prior to implementation but little assistance was given when they were working out 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., in press, a) and then designed action research projects to use the data they collected to improve their own practices (Ross et al., in press, b). In all phases, including implementation in the classroom, there was intensive interaction between the future mentors and three academics. The CLEAR mentors had much more time than the action research teachers in this study to talk with their peers about assessment and instruction.

A second limitation of the study is that it does not provide data on each step in a causal chain linking in-service to student achievement. Such a chain would demonstrate that (i) in-service influences (ii) teacher cognitions, which influence (iii) teachers' assessment practices, which influence (iv) students' self-evaluation accuracy, (v) cognitions about assessment, (vi) student

motivation and (vii) achievement. The study reported here links (i) and (v). In our previous research we linked (i), (ii) and (iii) by tracking a small sample of teacher-researchers as they reconceptualized assessment and re-invented their practice. In our current research we are measuring the impact of a modified form of the action research treatment described in this article on various student outcomes, linking (i) to (iv), (v), and (vi). Preliminary results indicate that teacher in-service increases students' accuracy in self-evaluation (reducing overestimates of performance) and contributes to academic achievement (narrative writing skills).

We conclude our study with renewed optimism about action research as a vehicle for designing local improvement projects controlled by teachers and assisted by outsiders, in this case, academics. We learned three things from the study. First, action research had a greater impact than skills development in-service on student outcomes, even though we provided less research training support for teachers than we had in previous projects. Second, the variations among students in feelings and beliefs about self-evaluation were striking in both treatments yet we overtly addressed student cognitions in neither. Third, we learned that action research focused on root (as opposed to branch) beliefs, behaviors and relationships takes more time than we had allocated. Teachers needed more time to work out how to accommodate an innovation that involves sharing control of a core teacher function with their existing beliefs about teacher and learner roles. Students also needed more time to understand what self-evaluation is and how it relates to their learning, in addition to learning how to do it. Our next efforts are building on these findings by continuing to build in-service on an action research platform, with stronger support for teachers' ability to collect and analyze evidence, and with greater attention to how students process changes in classroom practice.

Endnote

¹ An earlier version was presented at the annual meeting of the American Educational Research Association, Chicago, March, 1997. The research was funded by the Ontario Ministry of Education and Training, the Social Sciences and Humanities Research Council of Canada, and the Durham Board of Education. The views expressed in the paper are not necessarily those of the Ministry, Council, or Board. The authors wish to thank Durham teachers who collaborated in the research: Barb Bower, Michelle Ferreira, Sharon Hopkins, Cheryl Hoyle, Anne Marie Laginski, and Dianne Serra. Administrative support was provided by Jim Craigen, Bev Freedman, Norm Green, Brian Greenway, and Don Real.

References

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

- Calhoun, E., & Allen, L. (1994, April). Results of school wide action research in the league of professional schools. Paper presented at the presented at the annual meeting of the American Educational Research Association, New Orleans.
- 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.
- 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.
- Dennis, M. (1994). Ethical and practical randomized field experiments. In J. Wholey, H. Hatry, & K. Newcomer (Eds.), Handbook of practical evaluation (pp. 155-197). San Francisco, CA: Jossey-Bass.
- 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.

- 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., & 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.
- 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.
- 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.
- 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., & 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.
- 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.
- Rolheiser, C. (Ed.) (1996). Self-evaluation: Helping students get better at it. Toronto: Visutronx.
- Ross, J. (1995,a). 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. (1995,b). Strategies for enhancing teachers' beliefs in their effectiveness: Research on a school improvement hypothesis. Teachers College Record, 97(2), 227-251.
- Ross, J. (1995, c). Assessment of group process and product. Final report of Ministry of Education Transfer Grant Project. Peterborough, ON: OISE Trent Valley Centre.
- 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. (in press, a). Student evaluation in cooperative learning: Teacher cognitions. Teachers and Teaching.
- Ross, J., Rolheiser, C., & Hogaboam-Gray, A. (in press, b). Teaching students how to self-evaluate when learning cooperatively: The effects of collaborative action research on teacher practice. Elementary School Journal.
- 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). Long term impact of a new approach to teacher education for primary science. 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.), Understanding teacher development (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.), Professional development in education: New paradigms and practices (pp. 135-150). New York: Teachers College Press.
- Wade, R. K. (1984). What makes a difference in inservice teacher education? A meta-analysis of research. Educational Leadership, 42, 48-59.