
Student Evaluation in Cooperative Learning: 
Teacher Cognitions

John A. Ross*
Carol Rolheiser
Anne Hogaboam-Gray

Ontario Institute for Studies in Education
of the University of Toronto

June 1997

*Corresponding Author:

OISE/UT Trent Valley Centre
Box 719, 150 O'Carroll Avenue
Peterborough, Ontario, Canada K9J 7A1
e-mail: jross@oise.utoronto.on.ca

CA\.
papers/coopevl3.doc
Student evaluation procedures used by Cooperative Learning [CL] teachers, and their feelings about them, have rarely been investigated. This interview study of 13 exemplary users of CL methods found that negative feelings ran through teachers' cognitions about evaluation: expressions of guilt, anxiety and uncertainty were frequent. There was a substantial gap between private and public knowledge. When individual insights were assembled in a composite picture, generic strategies emerged, for example, for teaching students their role in self-evaluation. In this study the mechanisms for making private knowledge public were weak. This deficiency was subsequently addressed by the teacher-researchers involved in this study in a series of action research projects.
Student Evaluation in Cooperative Learning:

Teacher Cognitions

Cooperative Learning [CL] demands changes of teachers and students, particularly in evaluation. Teachers who have evaluated students as isolated learners have to figure out how to evaluate students working in groups. Guidelines for giving students feedback on group work abound in CL manuals (e.g., Authors, 1991) but little is known about the effectiveness of these strategies, the frequency of their use, or whether teachers and students find them satisfactory. Pressure to reform student evaluation practices creates additional tensions for CL teachers. The purpose of our research was to investigate how exemplary implementers of CL approaches handled these tensions. We wanted to find out what they were doing, why they did it, and whether they felt they were successful. Our goal was to identify workable evaluation strategies that other teachers might try in their own classrooms.

Motivation for the Study

Teacher Attitudes to Assessment

Teachers view student evaluation as an important teacher function and devote substantial classroom and teacher preparation time to it. Teachers rely on evaluation instruments of their own creation (Herman & Dorr-Bremme, 1983; Wilson, 1990), primarily because these measures are better integrated with their teaching practices than externally produced tests. Yet teachers give themselves low proficiency ratings on evaluation tasks and believe they could benefit from practical in-service (Gullickson, 1986; Impara, Plake, & Fagar, 1993; Marso & Pigge, 1992), even teachers who have experienced fairly lengthy training (e.g., 1-3 days in Bennett, Wragg, Carré & Carter, 1992). In addition, feedback from researchers about the quality of teachers' assessment practices is often negative. Anderson (1989), for example, criticized teachers' tests because few of their items measure higher order objectives and item writing guidelines are frequently violated.

Shift in Conceptions of Assessment

Conceptions of good assessment are moving toward direct observation of complex performance rather than brief written tests that correlate with the target aptitudes (Linn, Baker & Dunbar, 1991). In these performance
assessments students are observed working on complex tasks (e.g., Baron, 1990; Shavelson, Baxter & Pine, 1992) or dealing with real life problems (Raizen & Kaser, 1989). These instruments are often administered to groups of students because group work represents out-of-school performance better than individual production (Webb, Nemer & Chizhik, 1995). Such approaches to testing would seem to be ideal for CL classrooms.

Teachers' responses to alternate assessment have been mixed. Mandated alternate assessment programs produce teacher resistance due to schedule disruption, concerns about consistency, and doubts about the usefulness of the data (Howell, Bigelow & Evoy, 1993; Madaus & Kellaghan, 1993; Wilson, 1992; Worthen, 1993). Yet, when teachers have the freedom to choose, there is enthusiasm for alternate assessment (Bateson, 1994; Calfee & Perfumo, 1993).

Alternate conceptions of evaluation escalate demands on teachers. Alternate assessment must be transparent (Fredericksen & Collins, 1989), meaning that the criteria for appraisal, the population from which tasks are drawn, the scoring key and interpretive schemes must be visible to students, even when the teachers who devised these procedures have an imperfect grasp of them. Asking teachers to engage students in setting evaluation criteria (Bellanca & Berman, 1994; Garcia & Pearson, 1994) intensifies demands. Authentic assessment standards require precise specification of what will be measured, identification of multiple levels of attainment, and descriptions of opportunities to learn (Linn, 1994). The heightened concern with the moral dimension of evaluation (e.g., Wiggins, 1993), requires that teachers support due process, and allow students to be assessed at an appropriate level of difficulty and when ready.

Making such changes is not easy. Briscoe (1994) found that when beliefs about teaching and the constructivist learning theory implicit in alternate assessment conflicted, conventional test practices returned. In Briscoe’s study, conflict centred on one teacher’s theory of how assessment influenced learning. The teacher believed that regular monitoring based on unambiguous criteria, such as work completed, stimulated student productivity. For him, the motivational power of assessment resided in the fairness of objective procedures. When he tried to use performance assessment he felt that objectivity was lost. He had little confidence in the rules he developed for interpreting student responses, he believed his grades favoured students he liked, and he felt
assigning a single grade to all students in a group was unfair. Although he tried to resolve these conflicts, he eventually returned to multiple choice testing. Lorsbach, Tobin, Briscoe, and LaMaster (1992) observed two teachers for whom the purpose of assessment was control of students, tests emphasized knowledge reproduction, and work completion was a heavily weighted grading criterion. Shifting to assessments based on observations and interviews to accommodate experiments with constructivist teaching created conflicts for both teachers. One teacher resolved the conflict by redefining her metaphor of assessment from that of “fair judgment” to providing a “window into a student’s mind” (p. 309), thereby reconciling assessment with her new conception of teaching. The other teacher did not resolve the conflict. At the end of the study the tension between his constructivist approach to teaching and objectivist assessment practices continued. Other researchers have reported teacher misconceptions about specific alternate assessment techniques. Ruiz-Primo and Shavelson (1995) found over-inclusion: teachers thought performance assessment was anything that involved manipulation of concrete objects. Oosterhof (1995) found under-inclusion: teachers treated only formal tests as valid assessment procedures and included informal methods like observations and oral feedback only after probing.

**Student Assessment in Cooperative Learning**

Although the authentic assessment paradigm challenges all teachers, the impact is especially acute for CL teachers. First, 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. Second, CL involves teachers sharing control of assessment with students through procedures like peer- and self-evaluation. Such strategies open grading to a new source of bias, making it more difficult for teachers to defend the grades they assign. Third, most of the advice (and instruments) for assessing students when working in cooperative groups focuses on social skills. But accountability demands on teachers emphasize the measurement of cognitive growth. Teachers are therefore challenged to balance the cognition and social assessment demands of CL.

Advice from educational research on student assessment in CL is scanty. A few studies (Authors, 1995a; Huber & Eppler, 1990; Johnson, Johnson & Stanne, 1990; Johnson, Johnson, Stanne & Garibaldi, 1990) found that
specific evaluation procedures, such as structured peer review of group processes, have a positive effect on student achievement. But these studies combined assessment with other treatment elements, leaving consumers of the research uncertain about the relative contribution of the assessment component.

Other studies isolated specific evaluation practices suitable for CL. For example, Archer-Kath, Johnson, and Johnson (1995) found that peer evaluation made a greater contribution to students' social skills when individual performances were distinguished from group totals. Conway, Kember, Sivan, and Wu (1993) experimented with alternate forms of combining teacher judgments with peer evaluation of group projects, finding that two complex procedures were received equally well by students. Several researchers compared individual and collaborative test taking. Webb found that group discussions while completing a test over-estimated individual performance (Fall, Webb, & Wise, 1995; Webb, 1993). But she also discovered that group tests revealed information about students' concept understanding that was not accessible in individual testing (Webb et al., 1995). Billington (1994) found that having pupils agree on test answers increased student retention over individual test taking. Comparisons of collaborative and individual tests with adult learners have been mixed, with some finding that collaboration increased performance on a subsequent task (Lambiotte, Dansereau, Rocklin, Fletcher, Hythecker, Larsen & O'Donnell, 1987) and others finding no differences (Griffin, McCown, Quinn & Driscoll, 1994; Griffin, Griffin, Warkentin, Quinn, Driscoll & McCown, 1995).

These studies suggest that student assessment may be problematic for CL teachers. But only one previous study has examined CL teachers' cognitions about evaluation and that study (Briscoe, 1991) was limited to a single teacher who had just begun to experiment with CL methods. Since teachers with greater CL experience might be more successful in handling the assessment challenges of CL teaching, we focused on the student assessment practices of accomplished CL teachers. We wanted to find out what they did and how they felt about it. We wondered whether exemplary CL teachers would find evaluation problematic, as suggested by previous studies involving teachers who were experimenting with new approaches to instruction and assessment.

Method
The research was conducted by a school-university partnership. Three academics shared decision making with a district committee of five experienced CL teachers, two principals who made CL their top school priority, and the curriculum supervisor responsible for CL in-service. The district contained 109 schools (56,000 students) in a large geographic area in Ontario (Canada) spanning moderately sized urban centres (the largest was 60,000) and a large number of small communities and rural areas. The district had made CL a major curriculum priority and devoted substantial funds for a range of professional development sessions for teachers interested in CL methods.

Sample

The district committee in consultation with other district officials including school principals selected the sample. The prime criterion for teacher selection was very successful use of CL techniques over an extended period. All thirteen teachers who were nominated agreed to be interviewed. Three teachers taught Primary (grades 1-3), two taught Junior (grades 4-6), four were Intermediate (grades 7-8), and four were secondary (grades 9-12). Nine teachers were female and all had been teaching for ten years or more. Each teacher had attended a full range of district in-service sessions (e.g., 3-4 day summer institutes) and some had been involved in their delivery. These teachers were connected to CL networks outside their district. Many had attended CL conferences and some had participated in CL institutes offered in the United States.

Teachers had considerable autonomy within a curriculum framework established by the province, the district, and their schools. At the time of the study, the province was promoting an outcomes-based approach to education. Widely circulated curriculum documents identified a set of cross-disciplinary outcomes to be achieved by all students and specified performance norms in mathematics and language for the end of each division. The teaching strategies to accomplish these outcomes were not specified. Teachers selected methods based on their experience, reading of educational research, peer observation, and participation in professional development. The outcomes-based curriculum documents co-existed with earlier materials that prescribed grade and subject content. Some provincial officials said the new replaced the old; others told teachers that the new re-focused the old.
Mandatory provincial assessment programs supported the outcomes-based approach. All grade 9 students participated over a two-week period in a highly structured performance assessment in reading and writing. The results were reported by school. The province conducted reviews of other programs in which more traditional evaluation approaches were used (e.g., multiple choice items were administered to students at various grade levels). These were supplemented with reviews organized by the district and some district schools were selected for national and international testing. In none of these assessments were results for individual students or teachers released. The district had firm guidelines on reporting to parents.

The district vigorously recommended CL as an important addition to teachers' instructional repertoires and allocated a major portion of its resources to promoting CL use. For example, in the summer of 1988 the district began a summer institute program focused on CL. The initial institute was an eight-day in-residence program combining CL training with culture building and change management. The commitment continued in subsequent years and the number of institutes offered grew. In the summer of 1997, for example, half of the 20 three-day sessions focused on CL topics. The district’s commitment to staff development was honoured in the fall of 1996 with the international Bertelsmann Prize for school district innovation.

Data Sources

Each teacher was interviewed at his/her school twice over a two-month period for 60-90 minutes on each occasion using semi-structured interview guides. Teachers were interviewed by two members of the district team (alternating between questioner and recorder) or by one of the academics. The first interview had four sections. It began by asking teachers to describe their use of CL: how they became involved, the training sessions they attended, the CL model(s) they used and their feelings about CL. The second section focused on beliefs about evaluation: There were probes asking teachers to describe their student assessment practices in CL (e.g., what do you assess during a CL lesson? What instruments do you use? What works well? Does your method of assessment differ in a CL lesson than in other learning situations?). They were asked about specific issues (e.g., how do you judge individual success on a group product? How do you adjust your assessment strategies for exceptional students? How do you report CL experiences to parents? How do you record CL data?). There were probes about
Cooperative Learning

roles (e.g., who is involved in student assessment? How much do you share with students about their work will be assessed? How do you ensure consistency among teachers re evaluation in CL). There were also questions about when and how often teachers assess in CL. The third section of the interview asked for specific examples that illustrated the teacher’s evaluation principles. The final section invited teachers to address any issues not previously raised. Interviewers made detailed notes of the sessions. Interviewers prepared interpretive notes on their interviews. A set of themes was derived from these notes using analytic induction (scanning the data for categories and relationships among them), constant comparison within cases (e.g., comparing a teacher response to one question to responses to related questions), comparisons across cases (e.g., looking for similarities and differences between elementary and secondary teachers), and triangulation with artifacts (such as instruments) provided by interviewees.

The second interview began with the interviewer describing one aspect of the teacher’s assessment methods the interviewer particularly liked and then asked if the interviewee had additional thoughts about evaluation that had occurred to them since the initial interview. The interview then explored five themes which emerged from the first interview. These were (i) linking teaching and evaluation (e.g., was evaluation a factor when you selected your approach to CL? Do you feel any tension between your evaluation methods and your CL teaching?), (ii) evaluation criteria (whether a series of key words drawn from the first interview, such as fairness and rigor, could be applied to their evaluation methods), (iii) specific evaluation methods (interviewees indicated whether and how they used methods that involved peer evaluation, anecdotal records, rewards, portfolios, and others), (iv) reflections on evaluation (e.g., what are the next steps for you in developing your evaluation methods? Do you have a support group with whom you can discuss evaluation?). (v) special issues for Intermediate/Senior teachers (e.g., do you have different strategies for assessing in rotary classes than in non-rotary classes? How do you balance individual with group accountability?). The final section was unique to each interviewee and focused on ambiguities in responses to the first interview (e.g., in your last interview I heard you say that CL contributed to final grades in two ways...do you think students are aware of the direct and indirect effects of CL on their final grades?). Detailed interview notes were again compiled.
Each interview was tape recorded and transcribed verbatim. One interviewee declined to be recorded on both occasions and two tapes were lost in the second interview due to malfunctioning equipment.

**Analysis**

Two of the academics independently reviewed all the notes produced by the interviewers and minutes of interpretive meetings of the project committee. In creating categories we focused on (a) information about evaluation, (b) that occurred frequently, (c) that was emphasized by the teachers we talked to or (d) that addressed themes described in previous research on teacher cognitions or conversely (e) that addressed themes not considered in previous studies (e.g., evaluation in CL needs to be grounded in teacher-student trust). The categories developed independently by each academic were merged into a single list. The three academics independently coded ten pages of one transcript. Areas of ambiguity were clarified and new categories were added. The coding scheme (in the Appendix) had four main categories (feelings and cognitions about evaluation, evaluation principles, evaluation practices, and training issues) with subcategories and sub-subcategories within each. A manual defining and illustrating each of the categories in the coding scheme was used by pairs of district committee members to independently code 8 examples from one transcript. Discrepancies (of which there were few) were resolved through discussion.

Pairs of committee members began independently coding the transcripts and negotiating discrepancies. Due to constraints on practitioners' time 80% of the transcripts were ultimately coded by one academic. No within- or between-rater reliabilities were compiled. Interviewer notes were used to replace the missing transcripts. The coding unit was the utterance, defined as a complete thought relevant to a category. Most utterances consisted of at least one sentence (many were much longer) but some sentences were divided among two or more categories. All the transcript data were included in the coding. The transcripts and codes were entered into text analysis software (Drass, 1986), which sorted the data into the categories of the coding scheme. The appendix shows the frequency and percentage of each of the main codes for the first and second interviews.
Descriptive cross-case summaries were written for each coding category by the three academics and were revised following feedback from other committee members. From these summaries the full committee drew a series of assertions and evidence for each was assembled.

**Results**

The summaries provided a rich array of findings. Three themes emerged: the frequency of negative affect, the gap between private and public knowledge, and the weakness of mechanisms for making private knowledge public.

**Negative Affect**

Anxiety, guilt, and uncertainty permeated teachers' talk about evaluation. Teachers expressed great confidence in their ability to implement CL effectively. But their feelings of competence in using CL teaching techniques did not extend to evaluation: "my...knowledge about the whole art and science of teaching is further ahead than...my level of proficiency in evaluation." Expressions of negative feelings vastly overshadowed positive sentiments. The gap between ideals and practice weighed heavily for these teachers ("I personally didn't feel that I had yet done a good enough job as a teacher."). They felt they were not evaluating often enough, that they failed to balance individual and group accountability appropriately, that their appraisals were imprecise and unsystematic, and that they were using unreflective and unsophisticated procedures. "Sometimes giving an A, B, C, or D is the quickest way to do it without having to do any anecdotal and reflection, and any thinking in terms of growth." Some teachers were uncomfortable talking about evaluation ("I felt insecure") and we frequently heard, "I feel quite inadequate at evaluation; it's something I struggle with."

Negative feelings arose from a host of factors, including teachers' conceptions of student evaluation. There was a belief that rigorous evaluation excludes the intuition and subjectivity that often characterizes CL teaching. "That's probably an area that I'm not too good at. . .I'm not a very linear thinker and maybe not as systematic as I should be and I do a lot based on gut reactions and feelings and that sort of thing." They saw evaluation as distinct from and less important than teaching. They selected instructional approaches without regard to the place of evaluation in the CL model that they used. Assessment was loosely coupled with teaching; it was done after the
important issues had been decided. These teachers frequently equated evaluation with record keeping ("getting [it] down in a book and having it all succinctly in one spot") and felt guilty that they were not taking as many notes as they thought they should ("keeping records is. . .mind-boggling. . .and then what do you do with all the data?").

Some teacher uncertainty was based on their perception that CL theorists contradict one another on assessment issues. "I know Kagan, and Wilson, and DeChamps make you almost swear not to give group grades. . .and I know the Johnsons feel slightly different about that." Additionally, recommended evaluation strategies conflicted with other goals they held dear: "I hesitate [with peer evaluation] sometimes, I think because we work so much on team-building and having people feel good about themselves. I hate to put anybody in a critical mode with their peers."

Teachers took responsibility for their perceived evaluation failings. Teachers attributed lack of success (e.g., when peer appraisal turned into a popularity contest) to teachers not spending the necessary time to teach students the proper techniques: "that's something that takes a lot of practice and maybe I haven't given my kids enough practice at it." The uncertainty teachers felt about appraising their own work translated into uncertainty about self-evaluation's effectiveness when used by students. "It's pretty personal and I'm not sure. . .a lot of teachers . . .know how to evaluate. . .is that something we should be putting on the kids when they maybe don't know all the parameters?"

Teachers held themselves accountable to others for their evaluation practices ("you have to address parents. . .you have to be ready and willing to justify what you're doing, to them and to the kids"). Some resorted to traditional assessments to bolster their claims ("I want proof. . .because I want to be able to say, when I write the report card, I know they can do this, and I want to be able to show it to the parents."). Some saw benefits in the accountability demands of their district ("it often makes us stop and look at what we're doing and think about it... Am I writing airy-fairy garbage on the report card?"), but still felt uneasy ("it's kind of scary sometimes").

As teachers became more experienced in CL they reported that their comfort with evaluation diminished. Teachers wistfully spoke of a dream in which the evaluation of students in CL was linear and ordered. "What you'd like to do is just keep a little file box with a card on [each student]. . .I tried that one year. . .HUH!"
increased as the layers of complexity became visible. "It's really complex. The way I do it now isn't the way I did it when I started with cooperative learning. Every time I get another year into it, I realize how much more I don't know."

Teachers wanted to improve their skills. Teachers felt there had been no professional development opportunities focused on student evaluation in CL contexts. Furthermore they identified a host of specific needs. Especially dissatisfying were deficiencies in diagnostic skills. "We're still not very good at identifying exactly in a concrete way where the learner is on whatever my goal is when I get them, where are they now a month and a half later, and where are they at the end?" Although stated as a problem for all students, teachers were particularly frank about their difficulty in assessing students with exceptionalities. Other specific needs were identified by individuals but the shared theme was "I wish I knew what evaluation was all about."

Private and Public Knowledge

The potential knowledge of the group exceeded the private knowledge of individuals. Knowledge about evaluation was distributed unevenly within the sample and no single individual possessed a definitive repertoire of strategies. Assembling individual insights into a composite picture revealed that these teachers could learn much from one another. Two examples are given: teaching self-evaluation and building trust.

When teachers' responses were collected a generic strategy for training students in self-evaluation emerged. Self-evaluation is a potentially powerful technique because of its impact on student performance through enhanced self-efficacy and increased intrinsic motivation. Evidence about the positive effect of self-evaluation on student performance is particularly convincing for difficult tasks (Arter, Spandel, Culham & Pollard, 1994; Maehr & Stallings, 1972), especially in academically oriented schools (Hughes, Sullivan, & Mosley, 1985) and among high need pupils (Henry, 1994).

Student self-evaluation was problematic for these teachers as it is for others (e.g., Cox, 1995 found that while a majority of teachers believed they should include self-evaluations in student grades, only 20% did so). The process was introduced later in the year because students found the process hard to learn ("I don't think they're ready to self-evaluate right off the bat because it is hard to put into words") and teachers saw it as "a very, very
hard thing to teach.” There was also uneasiness about how parents might view self-evaluation (“whether it was because I was afraid of backlash from parents or I just didn't feel comfortable enough, I wanted to play with this a long time”). Moreover students had relatively little exposure to self-evaluation prior to entering the classes of our teachers and for a few teachers it was an infrequently used tool (“I don't do that a lot.”).

By assembling strategies used by individual teachers, a four step procedure for training students in self-evaluation was developed.

(i) Involve students in defining the criteria that will be used to judge their performance. Teachers believed that student participation in the identification of evaluation criteria increased motivation ("if you don't spend a lot of time on evaluation and you don't include kids, they kind of get turned off") and appealed to their sense of justice ("students think it's neat or fair, when they have a part in it"). The older the student, the more likely they were to be invited to contribute.

(ii) Teach students how to apply the criteria to their own work. The basic technique was teacher modelling, especially with younger children. "I...teach them how to frame what they're going to say and then as time goes on it becomes part of their language.” Many examples were provided of what particular categories meant, using language that connected criteria to evidence in the appraisal. For example, in a 1-5 scale measuring active involvement in learning: "Would you put yourself at a 1 because you were getting up and sharpening your pencil or going to the washroom or a 5 because even if I was spitting nails out here at the front you wouldn't notice?"

(iii) Give students feedback on their self-evaluations. In every grade teachers introduced other data as a check against bias in self-evaluation. The source of correlative evidence in the elementary panel was teacher appraisal ("they evaluate themselves and then we evaluate them"). The data were pooled in a student-teacher conference. If the discrepancy between the two assessments was too great ("if we're too black and white") the teacher ultimately decided ("I'm the one who settles it") but the decision was not arbitrary ("they really get a chance to make their case"). Such comparisons were reported to be fair because teachers and students applied the same
Cooperative Learning

criteria and in some instances they used the same instruments ("the same sheet transfers to a self-evaluation sheet that the kids use").

Peer assessments were used to balance self-evaluations of older children. Intermediate and Senior students' self-appraisals had to be confirmed by students who had observed the work that was done. "Before a person can let his mark for himself stand, all of the group members should agree on that. . .he makes a judgment on his performance that is validated by the group." If the discrepancy between the individual and his/her peers was too large the teacher mediated. "If a person gives themselves really high individual marks and. . .the peers have marked them low, we conference that." In contrast, younger children did not negotiate with their peers, although some teachers encouraged students to declare their self-evaluations ("they shared round robin what they gave themselves and why").

Older students were required to justify their self-evaluations. With younger students the goal in student-teacher conferences was to reach consensus on a single interpretation. Teachers of older students were more willing to retain the integrity of the student's self-evaluation but introduced some controls when calculating the contribution of self-evaluations to students' final grades. This could take the form of averaging the self-evaluation with the teacher's judgment ("I . . .averaged out their mark and my mark so that it was somewhere in between the both of us."). Other teachers included the self-evaluation as a limited portion (e.g., 15%) of the final grade. Teachers who translated self-evaluations directly into grades insisted students justify the grade they assigned themselves with an explanatory comment ("students were aware that. . .I needed the comment, the reflection"). In contrast younger students were allowed to give a simple appraisal of their work without elaboration and were required to explain their judgment only if it differed substantially from the teacher's judgment.³

(iv) Help students develop productive action plans. The primary uses of self-evaluation data were for reporting to parents and for program planning. A few teachers used self-evaluation data to encourage student reflectiveness about their behavior and identify how it could be changed.

It took a number of times and finally it became reflective. . .they looked at their mark, then looked at how they were spending their time, and realized that they weren't spending their time wisely, and therefore,
their mark suffered. So, when they made that correlation, that seemed to make them focus a lot more on process. So, that was a good learning experience for all.

Overall, most teachers neglected this stage, providing little guidance to students.

The second example concerns building trust. All teachers agreed that trust, between teacher-student, among students, between teacher and community, was essential to effective evaluation. Each teacher had at least one strategy for building trust but no one had as complete a list as the group as a whole. The strategies they used included: (a) Establish norms that legitimate discussion of grading (“I always tell students to approach me or any teacher if they’re unsure of their marks or how something was marked.”) (b) Provide direct training in how to talk about it (“I explain to them how to ask a teacher without the teacher being put on the defensive.”) (c) Ask for student trust directly (“I’ve said to them ‘don’t judge me too harshly on this: I’ve only been playing with it [evaluation in CL] for about two years.”) (d) Discourage students from blaming a group member when the group mark is lower than expected (“This was not a time to put anybody down; it had to be a group effort.”) (e) Teach students how to negotiate discrepancies between peer- and self-evaluations (“disagree in an agreeable way, with the ideas not the person”).

Making Private Knowledge Public

Mechanisms enabling teachers to learn from one another about evaluation were weak. Many teachers reported that there was no one with whom they could share professional thoughts. A few teachers spoke glowingly about the benefits of teacher collaboration in their department or school. Working with other teachers reduced duplication of effort (“why reinvent the wheel”); produced new ideas (“every time somebody has a good idea we share; you can’t do it all yourself”) and resources (“our files are all open to each other”); helped avoid problems when implementing new strategies (“if there were pitfalls, and there always are. . .you don’t have to do it over the same way”); identified areas of common need (“a support group at least makes us aware we’re all having difficulty in the same area”); communicated enthusiasm (“somebody will come back and they’ll be excited and in turn will make five other people excited about what they are doing”); and helped teachers maintain their motivation (“it’s just the feeling that we’re in this together”). These conversations focused on the exchange of teaching ideas. Very
little sharing of evaluation techniques was reported. "When you're in teachers' college you do a bit of evaluation stuff but . . . we haven't had any updates on anything new in evaluation for a long time."

Teachers who did not have a support group wished they did. "I'd really like to see grade level meetings that discuss things like evaluation." Even teachers who had a nonteaching colleague to interact with on assessment ("the academic resources teacher . . . and I spend a lot of time talking about evaluation") wanted more contact with classroom teachers on evaluation issues. These teachers reported feeling uncertain about standards ("where am I? . . . where are my kids on that continuum?") and posed fundamental questions about evaluation practice that required a collective response.

I wonder about how evaluation differs from one teacher to another, one area to another. Do we need [common] standards or do we evaluate to the needs of our school and students? If so, what happens when a student moves?

Discussion

This study produced three main findings. The first finding was that teachers expressed guilt, anxiety, and uncertainty about their methods for evaluating student work. One reason was that teachers' beliefs about assessment were not shifting as rapidly as their practices. Many of these teachers defined evaluation as the precise recording and aggregation of objective student scores. Yet the data they were collecting were laden with subjectivity. Teacher's intuitions, individual student bias, and peer sentiment ran through the evaluations. Although the teachers were beginning to use alternate assessment procedures, they had not made the epistemological change required to embrace performance assessments with confidence. They had difficulty recognizing that the transparent, democratic, and subjective evaluation appropriate to CL was a valid form of appraisal. Some teachers, as in Briscoe (1994), responded by reverting to traditional assessment methods to meet accountability demands. In addition, these teachers were experiencing an intensification of work, a chronic sense that the pace of production was being speeded up (Apple, 1986). They needed time to reflect but they were so caught up in new initiatives that the connections between one innovation and the next were murky. They treated each innovation as a separate entity rather than a new episode in the continuing process of change.
Our second finding was that there were few mechanisms for sharing information about evaluation. Teachers had to create a new synthesis of teaching and testing on their own. The stable synthesis of assessment and instruction in transmissive teaching was coming apart. Provincial and district authorities were constructing a curriculum framework for each of the pieces (performance assessments replacing decontextualized tests, peer learning strategies pushing out recitation of teacher and text) but the burden of fitting the pieces together rested with teachers. District in-service gave little attention to tailoring evaluation practices for the CL classroom. Limited interaction with colleagues on evaluation procedures inhibited the construction of private and public standards that CL teachers could use to judge their work.

Although a modest amount of teacher guilt may inspire a search for new strategies (Hargreaves, 1994), it was not happening here. Our teachers talked about improving their evaluation techniques, but few attempts to initiate the process were reported, in contrast with their vigorous efforts to acquire new instructional techniques. Teacher efficacy (i.e., teachers' confidence in their ability to perform the actions required of a professional teacher) might be the intervening factor. Individuals who see themselves as professionally competent adopt challenging goals, try harder to achieve them, persist despite setbacks, and develop coping mechanisms for managing their emotional states (Bandura, 1993). Teachers with low efficacy beliefs avoid situations that expose their ability. Past research (reviewed by Authors, 1995, c) has treated teacher efficacy as a homogeneous trait: confidence in one dimension of teaching predicts confidence in another. But recent studies (e.g., Authors, 1996, a) show differences in efficacy beliefs within-teachers, suggesting that teachers would have no trouble maintaining high confidence in their future performance in most teaching functions, even as they harbored doubts about their effectiveness in handling evaluation tasks. If so, negative affect about evaluating students in CL is likely to be an inhibiting factor depressing the willingness of teachers to learn more about evaluation and decreasing the likelihood they will be successful in meeting new evaluation needs.

Our third finding was that what teachers knew about evaluation as a group exceeded their individual knowledge. When assembled, this composite knowledge could enrich the assessment practices of other CL teachers. But how might this collective knowledge be delivered? The study reported here was conducted by a
partnership of academics and practitioners (teachers and administrators) in which research tasks were shared. As this study unfolded the five experienced CL teachers on the research team reported some of the same feelings of uncertainty about their own evaluation methods. But as the study progressed these teacher-researchers became more knowledgeable about the strengths of their practice, more precise about what they wanted to change, and more confident about their ability to do so. At the end of the study reported here each teacher conducted an action research project to design and test new strategies for evaluating students' CL. Each of these action research projects had a unique focus within a common theme of teaching students how to evaluate their work. The results (Authors, 1996, b) indicated that these teachers' became more confident about the defensibility of their evaluation practices.

The experience of the teacher-researchers in using the knowledge they acquired by interviewing 13 exemplary users suggests that the knowledge base of teachers can be expanded through collaborative action research. It suggests that the development of teachers' capacities, individually and collectively, is a process of reflective inquiry that is best-addressed in collaborative teams of teachers and academics (Authors, 1993; 1995, b). But these strategies will be more successful if pursued in concert with other steps that influence the ability of teachers to integrate multiple innovations. These include the development of collaborative structures within schools to enhance instructional decision making and induction of new teachers, shared decision making on school-wide issues that influence teaching and assessment, and political action to influence the curriculum context in which classroom practice is enacted.
REFERENCES


DRASS, K. (1986) Tap 1.0: text analysis package (Dallas, TX: Mimeo).


Footnotes

1 Removed for blind review.

2 Some researchers distinguish between evaluation and assessment in order to differentiate the data collection from data interpretation. But since there is no consistent pattern of usage, the terms evaluation, assessment and appraisal will be used interchangeably in this article.

3 Student teacher conferences were negotiations. If a student provided evidence about student performance that was not visible to the teacher (e.g., because she was in another part of the room when the evidence was generated), the student view prevailed. If the student provided no additional evidence the teacher's view prevailed. Teachers worried that over-ruling students would have a negative effect on their willingness to self-evaluate but teachers were more concerned that students apply the performance criteria in the way teachers intended.
### APPENDIX

**CATEGORIES FOR GROUPING DATA**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1. <strong>Feelings and Cognitions About Evaluation</strong> (1&lt;sup&gt;st&lt;/sup&gt; interview: ( N = 122, 16% ); 2&lt;sup&gt;nd&lt;/sup&gt;: ( N = 109, 18% ))</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>i)</td>
<td>Negative feelings about evaluation (e.g., anxiety, guilt, uncertainty)</td>
</tr>
<tr>
<td>F</td>
<td>ii)</td>
<td>Positive feelings about evaluation</td>
</tr>
<tr>
<td>F</td>
<td>iii)</td>
<td>Cognition</td>
</tr>
<tr>
<td>F</td>
<td>iv)</td>
<td>Other feelings and cognitions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2. <strong>Evaluation Principles</strong> [1&lt;sup&gt;st&lt;/sup&gt; interview: ( N = 309, 41% ); 2&lt;sup&gt;nd&lt;/sup&gt;: ( N = 223, 37% )]</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>i)</td>
<td>Characteristics of effective evaluation in CL:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• grounded in student/teacher and student/student trust</td>
</tr>
<tr>
<td>N</td>
<td>ii)</td>
<td>• explicit criteria</td>
</tr>
<tr>
<td>N</td>
<td>iii)</td>
<td>• involvement of kids in generating criteria and interpreting it</td>
</tr>
<tr>
<td>N</td>
<td>iv)</td>
<td>• &quot;fair&quot;</td>
</tr>
<tr>
<td>N</td>
<td>v)</td>
<td>• includes individual and group accountability</td>
</tr>
<tr>
<td>N</td>
<td>vi)</td>
<td>• regular and ongoing</td>
</tr>
<tr>
<td>N</td>
<td>vii)</td>
<td>• immediate feedback</td>
</tr>
<tr>
<td>N</td>
<td>viii)</td>
<td>• variety of instruments and methods</td>
</tr>
<tr>
<td>N</td>
<td>ix)</td>
<td>• self-evaluation follows practice with other forms of evaluation</td>
</tr>
</tbody>
</table>
Cooperative Learning

Nd ii) Use of data

Ndk • for feedback to kids

Nd p • for feedback to parents

Ndt • for teacher planning and decision making

No iii) Other evaluation principles (e.g., rigor/consistency)


Cd i) Divisional differences

Cdy • younger kids get more observation and social skills

Cdo • older kids get more peer evaluation and extrinsic rewards

Cdm • good strategies are modified by age

Ca ii) Differences in assessing cognitive and social outcomes

Cas • social skills are recorded and reported more anecdotally

Cad • different methods for different outcomes

Cg iii) Grading issues

Cgd • dissatisfaction with common group grade

Cgp • problems aggregating social & cognitive performance in a grade

Ce iv) Exceptionalities

Cel • lack of concrete strategies

Cm v) Missing links

Cmp • portfolios

Cmt • cooperative test taking

Cmr • cooperative test review

Co vi) Other evaluation practices (e.g., evaluation in rotary timetables)
4. Training Issues [1st interview: \(N=22, 3\%\); 2nd: \(N=57, 10\%\)]

Ts  
  i) Strategies for training students in evaluation methods (training need for real participation)

Tp  
  ii) Need to educate parents about evaluation issues; parental preference for more traditional norm-referenced, paper and pencil approaches

Tn  
  iii) Perceived teacher learning needs (interviewee or other teachers)

Tm  
  iv) Support mechanisms for learning (e.g., teacher collaboration)

Td  
  v) Lack of discourse

To  
  vi) Other training issues

5. Other Issues [1st interview: \(N=142, 19\%\); 2nd: \(N=73, 12\%\)]

OTm  
  i) Training

OIn  
  ii) Integration of other dimensions of teaching with CL

Note: A negative sign can be placed after any of the codes.
An earlier version of this article was presented at the annual meeting of the Canadian Society for the Study of Education in Montreal, June 1995. The research was funded by the Social Sciences and Humanities Council of Canada, the Ontario Ministry of Education and Training, and the Durham Board of Education. The views expressed in this paper are not necessarily those of the Council, the Ministry or the Board. The authors would particularly like to thank the teachers and administrators in the Durham Board of Education who contributed their insights to the issues reported here and Carolyn Brioux who transcribed the interviews.

References


