
Influences on Student Cognitions About Evaluation

John A. Ross*

Carol Rolheiser

Anne Hogaboam-Gray

Ontario Institute for Studies in Education

University of Toronto

September, 2001

*corresponding author

Dr. John A. Ross
Professor & Field Centre Head
OISE/UT Trent Valley Centre
Box 719, 633 Monaghan Rd S.
Peterborough, ON
K9J 7A1
CANADA
Author Biographies

John A. Ross is Professor of Curriculum, Teaching, and Learning at the Ontario Institute for Studies of Education at the University of Toronto and head of the Institute's field centre in Peterborough, Ontario. His research interests are student evaluation, program evaluation, and instructional development.

Carol Rolheiser is Associate Dean of Education and Associate Professor of Curriculum, Teaching, and Learning at OISE/UT. Her research interests are teacher development and school improvement, instructional and evaluation innovation, and teacher education reform.

Anne Hogaboam-Gray is the senior research officer at the OISE/UT field centre in Peterborough and a doctoral candidate in the Department of Curriculum, Teaching, and Learning, OISE/UT. Her research interests are qualitative and quantitative data analysis procedures and student evaluation.
Abstract

Interviews on two occasions with 71 grade 2, 4, and 6 students in a multi-ethnic setting in Toronto, Canada found that student cognitions about evaluation mediated the relationship between evaluation and achievement. Parents, peers, and student characteristics influenced student cognitions about evaluation. Parents identified the evaluation dimensions their children should attend to, raised student aspirations, stated how well student work attained standards, and recommended actions children should take in response to the evaluation. Peer interpretations influenced whether a given performance was viewed as superior or inferior. Older student peers focused attention, to a greater extent than parents, on specific aspects of student performance that could be ameliorated through self-remediation. Children became more sophisticated evaluation consumers as they grew older. Females processed evaluation data more productively than males. There were few cultural differences in response to evaluation. Students responded to traditional and alternate evaluation in very similar ways.
Influences on Student Cognitions About Evaluation

Student evaluation data are collected for a variety of purposes: accountability, student placement, program evaluation, accreditation, and to provide students with information they can use to improve their learning and performance. It is the last purpose that is our concern. In this study we interviewed a sample of grade 2, 4, and 6 students to deepen our understanding of how they process evaluation information regarding their daily work. Our central goal was to find out whether student cognitions about evaluation were influenced by parents and peers and by student background characteristics such as age, gender, and ethnicity. We also examined whether student processing of evaluation data was the same for alternative assessments (in this study self-evaluation) as for traditional evaluations by the teacher. (In this article we define evaluation as any appraisal of a student’s work. We reserve the term assessment for particular evaluation types, as in authentic, performance, collaborative, peer or portfolio assessment.)

Literature Review

Previous Research

There is a substantial body of research investigating the effects of teacher feedback on student achievement. This research has demonstrated that evaluation by a teacher contributes to higher achievement. For example, Kluger and DeNisi (1996) found a positive effect ($ES=.40$) in their meta-analysis of 131 studies. Bangert-Drowns, Kulik, and Kulik (1991), in an earlier meta-analysis of 40 studies, found a moderately strong ($ES=.54$) positive relationship between the frequency of evaluation and student achievement. But both review teams found substantial variation. There were a surprisingly large number of studies reporting that evaluation had negative effects on student performance.
Black and William (1998), in a review of 250 investigations of evaluation and classroom learning, identified a broad range of variables that modify the relationship between evaluation and achievement. Among the most compelling are teacher evaluation practices that have unintended negative effects on student motivation. For example, Torrance and Pryor (1998) provided a finely grained analysis of teacher-student evaluative dialogues. They found that the teacher’s actions were not informed by a theoretical understanding of motivation research and that this deficiency led to inappropriate use of extrinsic rewards, an emphasis on performance rather than learning goals, attributions for failure to external factors beyond the control of the child such as task difficulty, luck or low-ability. These feedback practices impeded the development of productive student goal orientations and efficacious attributions (such as attributing success to ability and failure to lack of effort). These teacher behaviours persuaded Sadler (1989) that students frequently do not benefit from teacher feedback even when it is valid and reliable. In Sadler’s account the heart of formative evaluation is communication of standards and strategies for self-improvement that would enable students to move from teacher- to self-monitoring. For this transition to occur, Sadler argued that students needed to understand the standards used to appraise their work, be able to use the standards to assess their performance, and have strategies for modifying their work during the act of production.

The attention given by these researchers to the social context of evaluation advances our understanding considerably, moving beyond treating evaluation as a one-sided process controlled by the teacher to a consideration of teacher-student dynamics. In our study we maintained the interest in how teacher actions impact on student cognitions but we expanded our treatment of the social context to focus explicitly on the role of peers and parents. We also maintained the previous interest in student characteristics (such as gender) on cognitions about evaluation.
Our Theoretical Framework

We assume that evaluation data do not influence achievement directly. In our view, the relationship between evaluation and achievement is mediated by student cognitions. Figure 1 displays our model of the processes. It is a model heavily influenced by social cognition theory (Bandura, 1997).

Figure 1 About Here

At the centre of Figure 1 are student cognitions about their achievement. The model highlights three processes that self-regulating students use to observe and interpret their behavior (Schunk, 1996). First, students produce self-observations, deliberately focusing on specific aspects of their performance relevant to their subjective standards of success. Second, students make self-judgments in which they determine how well their general and specific goals were met. Third are self-reactions, interpretations of the degree of goal attainment that express how satisfied students are with the result of their actions. This representation is comparable to the elements of the feedback system outlined by Black and William (1998), the difference being that we treat their fourth stage (a mechanism for reducing the gap between actual and desired performance) within the goal setting process described below.

Evaluation data provided by the teacher is intended to influence all three processes. Feedback influences the first process (self-observation) by directing student attention to particular performance dimensions, particularly if there is a shared understanding of criteria communicated through rubrics and exemplars (Arter, Spandel, Culham, & Pollard, 1994). Teacher feedback influences student judgments about the degree of their goal attainment (the second process), particularly if the teacher places student work on a continuum of achievement by comparing the work to overt standards (Ross, Rolheiser, & Hogaboam-Gray, 1999).
Evaluation data also influence degree of student satisfaction (the third process), particularly if feedback includes praise explicitly linked to the quality of the performance (Cameron & Pierce, 1994) but not if the praise is linked to effort without regard to the worth of the outcome (Lepper & Hodell, 1989). Evaluation information provided by the teacher is weighed against the observations, judgments and reactions of students generated during and after the performance. Formal evaluation data may complement or compete with these self-responses depending upon student perceptions of the credibility of the teacher’s evaluation and the child’s own.

Student processing of evaluation data may be influenced by parent and peer interpretations of the performance as well as by the teacher’s evaluation of it. For example, classmates’ achievement affects students’ perception of their achievement through social comparison. Students rate their performance lower if they are surrounded by high achievers (Marsh, Chessor, Craven, & Roche, 1995).

Student processing of evaluation data contributes to self-efficacy beliefs, i.e., student perceptions of their ability to perform the actions required by similar tasks subsequently encountered. Evaluation data may contribute to positive expectations about future performance through three processes: 1) by providing evidence that the student has attained a specific standard—the most compelling source of efficacy beliefs, 2) through vicarious experience (i.e., observing that individuals similar to the student were successful), and 3) through persuasion if the data comes with teacher claims that it demonstrates the student is able to be successful if he or she expends appropriate effort.

Self-efficacy beliefs in turn influence goals and effort. Students with greater confidence in their ability are more likely to visualize success than failure and they set higher standards of performance for themselves. Student expectations about future performance also influence effort.
Confident students persist. They are not depressed by failure but respond to setbacks with renewed effort. For example, students with high self-efficacy interpret a gap between aspiration and outcome as a stimulus while low self-efficacy students perceive such a gap as debilitating evidence that they are incapable of completing the task (Bandura, 1997). Goals and effort are also linked: Students are more likely to persist if they adopt goals that have unambiguous outcomes, that are achievable in the near future, and that are moderately difficult to achieve (Schunk, 1981). Self-efficacy also contributes to other positive motivations. Students with high self-efficacy are more likely to attribute success to their ability and failure to insufficient effort, to adopt learning rather than performance goals, and are more likely to be guided by intrinsic rather than extrinsic motivations (Bandura, 1997).

Positive self-evaluations play a key role in fostering an upward cycle of learning in which the child sets higher goals and commits more personal resources to learning tasks. Negative self-evaluations lead students to adopt social rather than academic self-images (Anderman, Anderman, & Griesinger, 1999), embrace goal orientations that conflict with learning, select personal goals that are unrealistic, adopt learning strategies which are ineffective, exert low effort, and make excuses for performance (Stipek, Recchia, & McClintic, 1992).

Studies of Student Cognitions About Evaluation

Almost all previous studies of student cognitions about evaluation focus on global attitudes toward evaluation and their impact on student motivation. Research on student processing of evaluation information is scanty.

Previous research identified several variables that moderate the relationship between evaluation and achievement. The strongest moderator is grade. Student cognitions about evaluation accumulate and have an enduring effect on subsequent encounters. Paris, Lawton,
Turner, and Roth (1991) administered multiple-choice surveys to students in grades 2-12 to identify strong developmental differences in student attitudes toward standardized tests used in mandated evaluations. They found a growing suspicion about the validity of test scores with older students believing that such tests did not measure the qualities of a good student and a growing conviction that parents and teachers did not care about the results. Older students decreased their effort on these evaluations, even as they became more anxious about the social comparison of scores. Paris et al. found that as students grew older two groups emerged: high achievers who took tests seriously and low achievers who did not. Paris, Turner, Muchmore, and Perry, (1995) found similar developmental patterns in grade 1-5 perceptions of portfolio assessments. Older students were more cynical about their value and less likely to identify advantages of their use. Paris et al. also observed grade-level differences in the criteria students used to select items to include in their portfolios and in their perception of the utility of portfolios for teachers and parents.

Several studies examined student attitudes to alternate assessments, anticipating that these would be less susceptible to the negative trends identified by Paris et al. (1991). Herman, Klein, and Wakai (1997) combined surveys with a small sample of interviews to find that grade 8 students responded differently to item types on mandated evaluations. Students described performance assessments as more interesting and challenging than multiple-choice items. Students claimed they tried harder on performance assessments because they had to explain more of their thinking than on traditional items. But students said they liked traditional items better than performance assessments because they thought traditional items were easier to understand and produced higher scores.
Most examinations of alternative assessment techniques examined student cognitions emerging from classroom use. Logan and Parkes (2000) found that grade 5 students liked traditional evaluation better than performance assessments because they believed only traditional items were valid measures. Performance assessments evoked confusion because the format was unfamiliar. Logan and Parkes’ findings may have been an artifact of study conditions; that is, performance assessments were an experiment that did not contribute to student grades. Morgan (1999) examined undergraduate student reflections on collaborative assessment (i.e., students pooled their answers and a common grade was awarded to all group members), finding that students were relaxed and confident with the procedure. Students may feel good about collaborative assessment in part because it over-estimates individual performance (Fall, Webb, & Wise, 1995), especially if the group is composed of higher ability students (Webb, Chizhik, Nemer, Sugrue, 1996). Ross, Rolheiser, and Hogaboam-Gray (1998) conducted focus groups of grade 2-12 students concerning their cognitions about self-evaluation (an approach in which students negotiate criteria, evidence, and grades with teachers; described in Rolheiser, 1996). We found that students liked self-evaluation better than evaluation by the teacher alone because it increased clarity about expectations, was fairer, and gave students feedback they could use to improve the quality of their work. Students reported that with teacher feedback they focused on the grade or areas they completed well, whereas with self-evaluation they focused on what they needed to improve.

There is limited evidence of gender effects. Broadfoot, James, McMeeking, Nuttall, and Stierer (1988, cited by Gipps, 1999) found that secondary school males were more likely than females to challenge evaluation decisions of teachers, a finding replicated with younger children.

Considerable debate has focused on whether alternate assessments are fairer to minority racial and ethnic groups than traditional tests. Traditional tests place a disproportionate number of minority students in lower tracks (Glaser & Silver, 1994), but the limited evidence to date suggests that the disadvantage of minority students in performance assessments is as great as on traditional tests (Bond, 1995). Garcia and Pearson (1994) suggested that alternate assessments might be better for cultural minorities because they allow teachers (in portfolio assessments) to include items that reflect minority performance better and can be tailored to focus on issues that are more relevant to minority students. These advantages are likely to accrue only if the teacher is knowledgeable about the minority groups in his/her classrooms. In addition, students from homes that emphasize adult authority may not see the benefits of such techniques as peer and collaborative assessment and may deprecate the value of self-evaluation.

In summary, previous research provides some evidence that student cognitions about evaluation are important mediators between evaluation and achievement, although the model depicted in Figure 1 has not been tested in its entirety. There were indications that students respond differently to alternative assessments than to traditional evaluation methods. The research also identified important moderators—grade, gender, ethnicity, and class—of the evaluation-achievement relationship. However, research to date has focused on global attitudes to evaluation with little examination of how data are processed by students. Finally, all the studies examined student cognitions about evaluation at a single moment in time.

Our study examined children’s cognitions about two types of evaluation information—traditional evaluation by the teacher and self-evaluation. We focused on self-evaluation as an
exemplar of alternative assessment because of our own interest in the method, because we were sure students would have experienced it (see below), and because we thought that students who were engaged in self-evaluation would be more articulate about how they processed evaluation data of any kind. We conducted the study across grades 2-6 because previous studies found that children’s attitudes were formed in these years and because alternative assessment is more frequently used in elementary than in secondary schools (Bol, Stephenson, O'Connell, & Nunnery, 1998). Most importantly we selected sites that had a high proportion of minority students.

Method

Sample

The study was conducted in six classrooms (two for each of grades 2, 4, and 6) in three school-university teacher education partnerships in Toronto, Canada. The teachers had been extensively involved in instructional innovation for several years and all were using a variety of student evaluation methods, including alternative assessment techniques. During the three months of the study six preservice candidates who had identified evaluation as a focus for their action research projects were assigned to the schools. The preservice candidates, in cooperation with the experienced teachers, developed rubrics with students, taught students how to use the rubrics to assess their work, gave students feedback on the accuracy of their evaluations, and helped students set goals (following the self-evaluation teaching model in Rolheiser, 1996). In previous research, these self-evaluation procedures contributed to higher student achievement (Ross et al., 1999; Ross, Hogaboam-Gray, & Rolheiser, 2001) and improved attitudes to evaluation (Ross et al., 1998).
Data were collected from 71 students, equally divided among three grades and two genders. Only 21 of these students spoke English at home but grade was confounded with home background: 69% of grade 2 students were English speakers but only 16% of grade 4 and 33% of grade 6. The remaining students represented 23 other language groups. The largest was Russian with 8 students, all in one grade 6 class.

Sources of Data and Analysis Procedures

Students were individually interviewed for 20-40 minutes on two occasions (February and April 1999). For the first interview students brought two self-selected evaluations; for the second they chose one self-evaluation and one evaluation by their teacher. They were asked to refer specifically to these evaluations when responding to questions about their feelings toward evaluation (e.g., “do you think evaluation helps you do better in school?”), their interpretations of evaluation data (“what did this evaluation tell you about how well you are doing in this subject?”), how they processed feedback (e.g., “did you talk to anyone about this evaluation?”), how they prepared for the evaluation (e.g., “what did you do to get ready for this evaluation?”), and a series of questions focused on self-evaluation (“what do you think counts as good work”).

Interviews were audio recorded. The day after each interview, interviewers compiled detailed interview notes while listening to the tapes. 50-60% of student comments were recorded verbatim with the remaining passages paraphrased. Analysis began by organizing data by interview question using QSR NUD*IST software. Themes were developed through constant comparison. We began with data from the second interview. We first established overall tendencies in student responses, searching for evidence of linkages between constructs in Figure 1. We examined differences in overall trends for gender and grade. We searched for factors influencing overall themes and the gender X overall and grade X overall interactions,
distinguishing ethnicity (English versus non-English mother tongues) and evaluation type (evaluation by the teacher and self-evaluation—there was insufficient information on peer- and performance assessment). The resulting themes were tested with data from the first interview using a process of reserving data for hypothesis confirmation, labeled referential adequacy by Lincoln and Guba (1985). The first round of interview data led to the modification of interpretations.

Credibility of the findings of the study was enhanced by 1) triangulating between data collection times and interpreters; 2) maintaining an audit trail by creating charts of relationships and counting instances (Miles & Hubberman, 1994); 3) searching for negative instances (Seale, 1999); 4) using member checks (draft themes were circulated to teachers for comment); and 5) recording data accurately (audio taping interviews).

Results

Assertion 1: Although students had mixed feelings about evaluation, they saw benefits in attending to the information evaluation provided.

Children expressed a variety of emotions about evaluation with positive comments outweighing negative ones. Regardless of their feelings, almost all children (67 of 71 interviewed) believed that evaluation helped them do better in school. Students reported that evaluative feedback told them what they were good at, what they needed to improve on, whether they needed to work harder, and clarified teacher expectations. Most students believed that the evaluation fairly appraised their work. They thought evaluation accurately predicted future performance, provided they prepared in the same way (e.g., “it means if I work hard again, I probably [will get] a good mark again.” [6-F-70-1] Most anticipated their subsequent performance would be higher because they would not repeat the mistakes revealed by the
evaluation. A few noted that subsequent tests would be harder “but we’re going to be learning more…so it doesn’t exactly tell me how I’ll do on the next test.” [4-F-49-1]

Virtually all students reported that they used evaluation data to plan future actions. Although some action plans were fairly general (“I will not talk to my friends when we’re on the carpet…and I will do well on my tests” [4-M-56-2]), others were quite specific. A grade 2 child, for example, planned to “write my printing neater and the picture a little bigger and neater and colorful.” [2-F-11-2] For a few students evaluation data were confirmatory. If they were pleased with their mark, they planned no changes. The vast majority of student comments reflected a deficit reduction theory of improvement. Evaluations identified errors that could be remedied or displayed an absence of a need for correction. No child suggested that evaluation revealed a strength that could be improved through formative feedback.

Assertion 2: Student judgments about their performance were informed by parental interpretations of evaluation data.

Almost all children shared evaluations with their parents, within limits. Students typically shared the evaluation (e.g., grade and teacher comment) without displaying the work on which the evaluation was based. Parental messages about the meanings of grades tended to be unambiguous even if the child’s probes were not. For example, a grade 4 child recounted, “I ask my dad if there is a person who has got this much mark, how good were they doing. He asks, ‘you got that much mark?’ And then I say ‘yeah’, and he says ‘you did pretty well’.” [4-F-53-1] Most students were more willing to show positive than negative appraisals. Several reported that they hid evidence of the latter in their desks at school. Given selective reporting, most children received positive parental feedback. For example, a grade 2 child’s parents “say wow and they hug me and kiss me.” [2-F-22-2] Others reported monetary or snack rewards. Conflict sometimes
ensued when parents were more concerned about the child’s marks than the child. Children reported that parents responded to poor evaluations with punishments (no TV), sadness, attempts to provide remedial help, and general injunctions to work harder: “study more and you'll get better.” [4-F-59-2] Not every parent was as concerned with academic accomplishments. For example, a grade 2 child reported that he felt good about his evaluation “because my mom and dad told me never to feel bad…because it doesn’t matter, it’s just a grade.” [2-M-09-1] Student reports on parental responses were similar for all types of evaluations.

Assertion 3 Older children processed evaluation data more thoughtfully than younger students.

The happiness about evaluation that prevailed in grade 2 was replaced in grade 4 with a broader range of emotions, many of which were negative. The importance of marks also increased with age with most of the change occurring from grade 2 to 4. The key age differences concerned how evaluation data were processed.

Students’ ability to make sense of evaluative messages increased with age. When asked what a particular evaluation was telling them, no grade 6 and fewer grade 4 than grade 2 students expressed uncertainty. Since uncertainty had a negative impact (e.g., “if I’m not sure I feel afraid” [6-M-80-1]), younger children sought parental assistance in interpreting their performance (e.g. “my mom and dad know a lot about it and they’ll tell me if I did good or not.” [2-F-11-1]) Older were more likely than younger children to state what they were doing well on, they identified needed improvements with increased precision (e.g., in grade 6: "It tells me that I know how to use a protractor and I know what kind of triangle it is." [6-M-89-2]), and action plans tended to be more specific (e.g., “I just need to improve on adding adverbs and adjectives." [6-M-71-2]) Grade 4 and 6 students were more likely than grade 2’s to indicate the basis for their interpretations, usually by citing scores they received. Grade 4 and 6 students were also more
likely to interpret high scores as an indicator of ability and low scores as an indication that more effort needed to be expended. For example, "[first evaluation:] it tells me that I'm doing well in music and that I'm good in music…[second evaluation:] I didn't do really well so I'll try harder. [4-M-37-2]

Older students drew on more resources to interpret evaluation feedback. Few grade 2 students shared evaluations with peers and sharing was limited to exchanges about marks. Some were advised not to share ("my dad says don't let anyone look at it cause it might make them jealous." [2-M-08-2]) For younger children, sharing of more than marks was interpreted as copying (e.g., "[I] share with some of my friends who have trouble…so they wouldn’t have to stay in for recess if they’re behind." [2-M-08-1]) In grade 4 most students shared with peers and all but one grade 6 student either sought peer comparisons or found they could not be avoided. Half the grade 4 peer comparisons focused on grades alone but in grade 6 almost all the students examined the work as well as the grade. Grade comparisons evoked negative behaviours, for example, "If you get a better mark than them, you will feel better." [6-M-61-2] Many students reported they shared only good results (e.g., "if I get a low mark, I don’t want anybody to know about it because I don’t feel good." [6-F-73-1] High scores sometimes evoked resentment: "They were kind of mad at me…This girl that sits beside, she got, in this test, perfect. They were mad at her too." [6-F-81-2] Perhaps for this reason, students in all grades preferred to share results only with their friends. Competitiveness increased with age, especially among boys. Some misrepresentations were reported, for example, "Sometimes I give them a high answer…because if I get it wrong I will look bad." [6-M-62-2] Most grade 6 students engaged in improvement-oriented comparisons that worked particularly well when peers had different strengths. For example, “Usually when I've had mistake on [one] they've got [it] right, and what I've got right
they've had a mistake on. We check each others work…I'd write the correct one in red so I know that I didn't answer it right.” [6-M-83-2]

Students who compared evaluations with peers were also more likely to compare their own work over time. The proportion of students making longitudinal comparisons increased from almost none in grade 2 to half in grade 4 to almost all in grade 6. Students were more likely to compare if the present evaluation differed from past scores. For example, “Like if I get a really bad mark or a really good mark then I would go and see why… I'd see what this story had that this story didn't have.” [6-F-70-2] Some students offered explanations for differences between the present and the past: “At the beginning of the year I didn't do so well because I didn't know what the teacher expected.” [6-M-71-2] Others focused on whether they were continuing to make the same mistakes as they had earlier. Students in grade 6 were more likely than grade 4 children to emphasize that comparisons were more credible when the assignments were similar. Only one of the grade 6 students was skeptical of the value of looking backward: “It doesn't make sense to me, because it's just showing that I did better this time…but it doesn't really help me to do better so I don't really compare them.” [6-M-87-2]

Older students continued to share their evaluations with parents but by grade 6 there was greater awareness of the potential for negative parental reactions. There was also more evidence of parental prompting for information. For example, "I showed this one to my mom because my mom wanted to know how I did on this one because I was writing for like a month." [6-F-70-2]

When assessing their own work, younger students focused on neatness and language conventions as key characteristics. For example, “if you can do it neat and proper sentences and it has to make sense and have proper punctuation.” [4-M-47-1] Neatness and language
conventions continued as important definers of quality work in grade 6 but they were embedded in a broader array of criteria than included rubric descriptions.

**Assertion 4** Females made more productive use of evaluation data than males.

There were no gender differences in overall attitudes to evaluation but how students interpreted the data was moderated by gender. There were gender-age interactions. Grade 2 girls tended to be more certain than grade 2 boys about what evaluation data meant. At all ages, girls offered richer interpretations than boys and provided more precise plans for acting on the data. Although only a few students reported that evaluation did not help them do better in school, all who did so were male. Girls drew on more sources of information than boys. In grade 2, girls were twice as likely as boys to report that they compared their current to previous evaluations. Girls who did not compare the present to the past gave a reason, such as no time in class. Among grade 4 children, girls were more likely than boys to compare results with peers. Yet males were much more competitive. For example, in grade 6, only males limited peer comparisons to marks, the least useful focus for discussion. Girls were also much more likely than males to report that rubrics were helpful when evaluating their work.

**Assertion 5** Students for whom English was not their mother tongue were more likely to be concerned with social comparisons.

There were few ethnic differences in student cognitions about evaluation. Non-English students (i.e., those who spoke a language other than English at home) were more likely to have positive attitudes toward evaluation. A much higher proportion of non-English than English students said they talked only about marks when comparing evaluations with their friends. In grade 6, where there were almost equal numbers of both groups, the ratio was 4:1. Non-English students were also more concerned than native English speakers with the consequences of such
comparisons. The six grade 6 students who said they compared evaluations only if they had good scores were all from homes in which English was not spoken.

**Assertion 6 Students made relatively few distinctions among types of evaluations.**

Students identified a number of positive features of self-evaluation: it gave them a sense of ownership, clarified teacher expectations, enabled them to participate in evaluation decisions with the teacher, and helped them set self-improvement goals. A few had a strong preference for self-evaluation. For example, a grade 6 student felt that student participation in defining evaluation criteria, a central feature of self-evaluation, should be incorporated into evaluation by the teacher: “instead of [the teacher] just deciding what the categories should be, we should make them up for tests, like agree on something.” [6-M-83-2] Others thought that teacher evaluation was more important because “That's the mark you are going to get on your report card…you better not think you are that good in the subject unless the teacher thinks you are that good.” [6-M-61-2]

When asked to respond separately to the examples of self-evaluation and evaluation by the teacher they brought to the second interview, students said virtually the same things about both types. The distinctions they made were mainly about the content of the evaluations. For example,

[In music, evaluated by the teacher] it told me that I was listening well and I followed instructions and I really focused on the task. [For homework, self-evaluated] it has to be just a little bit neater and I always have to get good at completing it and I need just a little more punctuation. [4-F-32-2]

When talking about evaluation by the teacher students indicated that the only thing that mattered was the teacher’s judgment. When they evaluated their own work they also considered
teacher expectations (females more than males) but in the context of a broader array of inputs. The most important of these were the level descriptions in rubrics (a joint student-teacher construction). For example, when asked how she gives herself a score, a grade 2 student replied, “I read the rubric and remember it and put the rubric in my head and if I’m starting my evaluation I take the rubric out and think about it.” [2-F-13-2] When asked what they compare their work to, students tried to ignore the performance of others when self-evaluating; for example, “I usually compare it to my own work because not other people’s marks are going on my report card, only my marks are going on it, so I need to see if I improved.”[4-F-52-2] Student conversations about self-evaluations were much less focused on marks than their talk about evaluations by the teacher, even though both types of evaluation contributed to report card grades.

Discussion

The study provided important information about how students process evaluation information (Figure 1) and the influences on their cognitions. The key finding concerns parental attempts to sway student interpretations. Parents identified which evaluation dimensions their children should attend to, usually the teacher-assigned mark. Parents tried to raise student aspirations and stated how well student work, according to its evaluation, attained standards. Parents also recommended specific actions children should take in response to the evaluation, such as exerting greater effort, and parents provided remedial assistance as preparation for subsequent assignments. Parental contributions to student cognitions strengthened student self-efficacy by encouraging students to believe that success was attributable to ability, failure was a consequence of lack of effort, and that students were capable of higher achievement on future tasks. Although some student resistance to parental influence was observed in the data, students
in every grade took parental input seriously. However, parental interpretations were almost invariably based on the evaluation alone, examined in isolation from the work itself. This result is worrisome in that parents may not be able to interpret the results of alternate assessments, a format they did not experience as children, without more support from educators than is presently provided. The result suggests a need for further investigation of parental influence, triangulating child and parent perspectives, especially for alternative evaluation approaches and in conditions of weak and strong support for parental interpretations.

Friends were more influential than other peers (Guay, Boivin, & Hodges, 1999 reported a similar finding). Peer interpretations mainly influenced the judgment aspect of student cognitions, affecting whether a given achievement was perceived as superior or inferior. In these classrooms, peers usually promoted high achievement, although accusations of excessive performance were reported. When focused exclusively on marks, social comparisons tended to be dysfunctional, leading to negative feelings and misrepresentations of achievement. Older student peers focused attention, to a greater extent than most parents, on specific aspects of student performance that could be ameliorated through self-remediation. This result suggests that peer assessment, which contributes to post-secondary achievement (Topping, 1998), might also benefit younger students.

The age trends that we found partly confirmed the increased negativity reported by Paris et al. (1991; 1995) but our main finding is that children became more sophisticated consumers of evaluation, as they grew older. They understood feedback better, drew on more resources to interpret it, and were more likely to report constructive action planning. We also found that females processed evaluation data more productively than males and that gender differences interacted with age in complex ways that warrant further inquiry.
We did not find large cultural differences in response to evaluation. Although encouraging, we are cautious about generalizing this result. We suspect that the distribution of our 71 students across 23 non-English language groups inhibited the detection of differences attributable to particular cultures. In addition, we conducted all interviews in English and avoided students with limited English.

We found few differences between traditional and alternate assessment. Students preferred self- to teacher-evaluation but we are reluctant to generalize this finding to all alternate assessments given studies reporting a preference for traditional over performance assessment (Herman et al., 1997; Logan & Parkes, 2000).

Although this study was not designed to test the full model in Figure 1, it provided ample evidence that student cognitions mediated between achievement and evaluation and that these cognitions were impressively influenced by parents, peers, and by student characteristics. The study also suggested that evaluation, at least in these classrooms, supported student learning and its impact could be strengthened if teachers attended more closely to parent-child conversations about evaluation. New directions in evaluation reform that focus on enhanced communication with parents (e.g., 3-way or student-led conferences) may be helpful in exploring this dimension further.
References


Figure 1: How student cognitions mediate the impact of assessment data on achievement.
End Notes

1 The research was funded by the Social Sciences and Humanities Research Council of Canada. The views expressed in the article are not necessarily those of the Council.

2 One case was lost when parental permission was withdrawn during the study.

3 The languages were: Arabic, Ashanti, Bengali, Bulgarian, Cantonese (3), Chinese, Farsi, Korean (3), Greek, Gujarati (4), Hindi, Indonesian, Japanese, Katchi, Latvian, Mandarin, Pashto, Punjabi (3), Somalian (4), Tamil (2), Urdu (6), and Vietnamese (3).

4 We avoided full verbatim transcription for reason of cost, following recommendations by Arksey & Knight (1999) and Stake (1995).

5 We followed this sequence: between grade 2 girls and grade 2 boys, grade 4 boys and grade 2 boys, grade 4 boys and grade 4 girls, grade 6 girls and grade 4 girls, grade 6 boys and grade 6 girls.

6 The code numbers display grade (2, 4, 6), gender (M or F), ID # (01-99), and interview (1 or 2).