The Reliability, Validity, and Utility of Self-Assessment

John A. Ross
University of Toronto

Despite widespread use of self-assessment, teachers have doubts about the value and accuracy of the technique. This article reviews research evidence on student self-assessment, finding that (1) self-assessment produces consistent results across items, tasks, and short time periods; (2) self-assessment provides information about student achievement that corresponds only in part to the information generated by teacher assessments; (3) self-assessment contributes to higher student achievement and improved behavior. The central finding of this review is that (4) the strengths of self-assessment can be enhanced through training students how to assess their work and each of the weaknesses of the approach (including inflation of grades) can be reduced through teacher action.

A large proportion of teachers (76% in Noonan & Duncan, 2005) reports using self-assessment at least part of the time, even though teachers express doubt about the value and accuracy of student self-appraisals. The doubts center on the concern that students may have inflated perceptions of their accomplishments and that they may be motivated by self-interest. Frequently heard is the claim that the "good kids" under-estimate their achievement while confused learners who do not know what successful performance requires, over-estimate their attainments. These concerns suggest, from a measurement perspective, that self-assessment introduces construct-irrelevant variance that threatens the validity of grading.

In this article I will examine research conducted on self-assessment for the purpose of addressing these practical questions posed by teachers:

1. Is self-assessment a reliable assessment technique?

2. Does self-assessment provide valid evidence about student performance?

3. Does self-assessment improve student performance?

4. Is self-assessment a useful student assessment technique?

Definitions

For the purpose of this article, I will follow Klenowski's (1995) definition of self-assessment as "the evaluation or judgment of 'the worth' of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes" (p. 146). This definition emphasizes the ameliorative potential of self-

assessment and focuses attention on its consequential validity. Although some of the research conducted on self-assessment has consisted of students appraising their work with little interpretative guidance, I will argue with Klenowski that the benefits of self-assessment are more likely to accrue when three conditions are met: teacher and students negotiate self-assessment criteria, teacher-student dialogue focuses on evidence for judgments, and self-assessments contribute to a grade (by students alone or in collaboration with teachers).

Although self-assessment has long been part of the repertoire of classroom teachers, assessment reform has increased its use. Key proponents of assessment reform (e.g.,Wiggins, 1993) recommend that students submit a self-assessment with every major assignment. Self-assessment is a valid instance of assessment reform (as defined by Aschbacher, 1991; Newman, 1997; Wiggins, 1993;1998) in that (i) students create something that requires higher level thinking (i.e., they interpret their performance using overt criteria); (ii) the task requires disciplined inquiry, (i.e., the criteria for appraisal are derived from a specific discipline); (iii) the assessment is transparent (i.e., procedures, criteria and standards are public); and (iv) the student has opportunities for feedback and revision during the task (e.g., by responding to discrepancies between the student's and teacher's judgment). Other important features of assessment reform, e.g., the extent to which the task represents real world applications of school knowledge, characterize some but not all self-assessments.

Some teachers find it helpful to distinguish between self evaluation (judgments that are used for grading) and self assessments (informal judgments about attainment) as suggested by Gregory, Cameron, and Davies (2000). Not everyone finds the distinction helpful; for example, the text on classroom assessment by McMillan (2004) uses the terms interchangeably. Throughout this article, I will use the term self-assessment to refer to both formative and summative data collections.

The term self-assessment is also used in the metacognition literature to refer to the judgments an individual makes on the basis of self-knowledge (Bransford, Brown, & Cocking, 1999). My review will focus on self-assessments conducted in classroom settings and will touch only briefly upon findings from lab investigations. For an extensive review of self-assessment in the context of metacognition, see Sundstrom (2005).

Why Teachers Use Self-Assessment

When asked why they include self-assessment in their student assessment repertoires, teachers give a variety of responses. (1) Most frequently heard is the claim that involving students in the assessment of their work, especially giving them opportunities to contribute to the criteria on which that work will be judged, increases student engagement in assessment tasks. (2) Closely related is the argument that self-assessment contributes to variety in assessment methods, a key factor in maintaining student interest and attention. (3) Other teachers argue that self-assessment has distinctive features that warrant its use. For example, self-assessment provides information that is not easily determined, such as how much effort students expended in preparing for the task. (4) Some teachers argue that self-assessment is a more cost-effective than other techniques. (5) Still others argue that students learn more when they know that they will share responsibility for the assessment of what they have learned.

Practical Questions Addressed by Researchers

Is self-assessment a reliable assessment technique?

Reliability, meaning the consistency of the scores produced by a measurement tool, can be determined in many ways. The internal consistency of self-assessments is typically high. For example, J. Ross, Rolheiser and Hogaboam-Gray (2002-b) had grade 5-6 students rate their performance on a 1-10 scale for each of five dimensions of mathematical problem solving. The internal consistency was .91. Similar results were obtained for grade 4-6 self-assessments in English (alpha=.84 in J. Ross, Rolheiser and Hogaboam-Gray, 1999). There is also evidence of consistency across tasks. Fitzgerald, Gruppen, and White (2000) examined the self-

assessments of medical students across two task formats: performance tasks (examination of standardized patients) and cognitive tasks (interpreting vignettes or test results). They found that students' self-assessments were consistent over a range of skills and tasks.

Less frequently examined is consistency between one time period and another. Blatchford (1997) found mixed evidence for long time periods. Blatchford reported that self-assessments were stable between ages 11 and 16 in mathematics, although not in English, a finding Blatchford attributed to feedback being less clear in English class than in mathematics. Blatchford found there was little agreement of self-assessments between ages 7 and 11 in either subject. There is greater reliability when the time periods are shorter. Sung, Chang, Chiou, and Hou (2005) had 14-15 year olds assess the quality of their web-designs on three occasions within a narrow time frame: after completing their designs, after viewing the designs of others in their own group and after viewing the best and worst designs in the class. Sung found no significant differences across occasions.

In summary, the evidence in support of the reliability of self-assessment is positive in terms of consistency across tasks, across items, and over short time periods. The studies showing adequate consistency involved students who had been trained in how to evaluate their work. There was less consistency over longer time periods, particularly involving younger children, and there were variations among subjects.

Does self-assessment provide valid evidence about student performance?

Validity in self-assessment typically means agreement with teacher judgments (considered to be the gold standard) or peer rankings (usually the mean of multiple judges which tend to be more accurate than the results from a single judge). Research on the self-assessments of university students produced mixed results. Boud and Falchikov (1989) reviewed 48 studies reporting self-teacher assessment agreement. In most, self-assessments agreed with teachers' ratings but the reviewers expressed concern about the quality of

many of the studies. There was extensive variation about what constituted agreement; the criteria used by teachers and students were frequently not defined; there were few replications involving comparable groups of students; some studies combined effort with achievement in a single rating; self-grading was not defined (e.g., it could be what a student deserves or what he or she expects to get). S. Ross (1998) also found mixed results for self-teacher agreement in studies of second language learning. He found a mean correlation of $r=.64$ ($N=60$ correlations) with wide variation among studies.

Student self-assessments are generally higher than teacher ratings, although exceptions have been reported (e.g., Aitchison, 1995 for middle school music students). Over-estimates are more likely to be found if the self-assessments contribute to the student's grade in a course (Boud & Falchikov, 1989). Young children may over-estimate because they lack the cognitive skills to integrate information about their abilities and are more vulnerable to wishful thinking. Butler (1990) found that self-teacher agreement increased from age 5 to 7 to 10 (correlations were $r=.16$, .38, .83 respectively). Agreement of teacher and student assessments is also higher when students have been taught how to assess their work (J. Ross et al., 1999; Sung et al., 2005), when students have knowledge of the content of the domain in which the task is embedded (Longhurst & Norton, 1997; S. Ross, 1998), when learners know that their self-assessments will be compared to peer or supervisor ratings (Fox & Dinur, 1988), and when the application of the assessment criteria involves low level inferences (Pakaslahti & Keltikangas-Järvinen, 2000 found greater teacher-self agreement for direct than for indirect measures of adolescent aggression).

Agreement of self-assessment with peer judgments is generally higher than self-teacher agreement (Bergee, 1997; McEnery & Blanchard, 1999). One explanation might be that students interpret assessment criteria differently than their teachers, for example, focusing on superficial features of the performance.

Fewer studies have examined other forms of validity, such as agreement with an objective criterion. Blatchford (1997) compared student self-assessments to standardized tests, finding that age moderated the relationship. Self-assessments were significantly correlated with achievement at age 16 but not at 7 years. A related literature examined the accuracy of recall of the results of standardized tests by university students. These studies found high correlations ($r = .87$ to .97) when students self-reported GPA and SAT scores when applying to graduate school (Cassady, 2001; Talento-Miller & Peyton, 2006); i.e., in conditions in which their self-reports could be easily checked against official documents. In broader settings, Kuncel, Credé and Thomas (2005) found that high achievers reported their college and high school grades accurately but accuracy was diminished for lower achievers, most likely by social desirability or self-enhancement factors. The correlations of self-assessments with external measures in the metacognition literature are mixed; for example, Ackerman, Beier and Bowen (2002) reported correlations ranging from $r = -.07$ to $r = .68$ across six subjects. Metacognition researchers found that the correlations of self-appraisals with external judgments were higher for younger than older children (Kaderavek, Gillam, Ukrainetz,, Justice, Eisenberg, 2004), for upper than middle or lower SES students (Pappas, Ginsburg, Jiang, 2003), and for boys than girls (Phillips & Zimmerman, 1990).

These studies suggest that self-assessments provide information about student achievement that corresponds only in part to the information generated by teacher assessments. The unexplained variation in self-teacher agreement has multiple sources, especially student inability to apply assessment criteria, interest bias, and the unreliability of teacher assessments. One systemic source of error might be that students include in their self-assessments information that is not available to the teacher, peers or standardized tests. As one student put it:

> *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 you worked at home or if you were just*

goofing off. (J. Ross, Rolheiser, & Hogaboam-Gray, 1998, p. 470)

In summary, the evidence about the concurrent validity of self-assessments is mixed. However, the review suggests that discrepancies between self-assessments and scores on other measures should be the stimulus for further inquiry, an invitation to review the evidence embedded in the learner's performance that might reveal student strengths and learning needs not addressed by the formal criteria.

Does self-assessment improve student performance?

Consequential validity is the argument that the worth of a test is determined by its consequences for students and others. For example, a valid assessment is one that contributes to student learning—if the assessment has a negative effect on student learning, the test is invalid (Moss, 1998). The inclusion of consequences as a dimension of test validity is a key element of student assessment reform.[1]

A few studies have demonstrated that asking students to assess their performance, without further training, contributes to higher self-efficacy, greater intrinsic motivation, and stronger achievement (Hughes, Sullivan, & Mosley, 1985; Schunk, 1996; Sparks, 1991). Other research found achievement outcomes in programs in which self-assessment was one of many treatment elements, although its unique contribution could not be isolated. For example, Fontana and Fernandez (1994) found large achievement benefits for mathematics students aged 8-14 in a program in which in which self-assessment was one of multiple strategies for increasing student control of learning.

Other studies have focused on the effects of training students how to assess their work. Although treatments vary, self-assessment training typically consists of systematic instruction in each of the elements that define self-assessment. For example, we applied strategies for teaching self-assessment in four stages: (i) involve students in defining assessment criteria (e.g., with teacher assistance construct a rubric that expresses

performance expectations and student criteria in language meaningful to students), (ii) teach students how to apply the criteria (e.g., model application of the rubric by assessing examples of performance), (iii) give students feedback on their self-assessments (e.g., engage students in evidence-based discussions of the differences between their self-assessments and assessments by peers or the teacher), and (iv) help students use assessment data to develop action plans (e.g., find trends in performance and identify short and long term strategies for overcoming weaknesses). Students trained in these processes over 8-12 week periods outperformed control samples in grade 4-6 narrative writing (J. Ross et al., 1999), grade 5-6 mathematics problem solving (J. Ross et al., 2002-b), and grade 11 geography (Ross & Starling, 2005). Effect sizes were of small to medium size; i.e., ES=.58 (for weaker writers), .40, and .50 respectively.

Less extensive training programs also produced positive results. In a review of six studies of secondary school student writing, Hillocks (1986) found that providing students with rating scales to assess their work improved writing quality. Arter, Spandel, Culham, and Pollard (1994) followed a similar strategy in which they gave grade 5 students scales to assess essay writing; treatment students outperformed controls (but on only one of the writing traits measured). Similarly Andrade and Boulay (2003) found that teaching grade 7-8 students how to use a rubric for self-assessment improved writing performance, although the effects were limited to females for one of two writing genres. McDonald and Boud (2003) found positive achievement effects for self assessment across a range of subjects.

Positive effects for self-assessment have also been reported for non-academic outcomes. J. Ross (1995) provided grade 7 students with transcripts of their conversations when working in cooperative groups and a coding scheme for interpreting the quality of their interactions. Students used the coding scheme to assess the frequency of help giving and help seeking in their groups. Self-assessment contributed to increases in positive interactions and a decline in off-task behavior. Henry (1994) developed a self-assessment tool for K-grade 1 students in which they compared what

they planned to do with what they did, using a check mark to indicate if actions matched plans. Henry found that use of the tool over 12 days contributed to higher student self-direction, but only for those with low self-direction on the pretest. Nelson, Smith, and Colvin (1995) devised a treatment in which disruptive grade 2 students were taught how to observe their playground behavior, make judgments about it, and obtain feedback from an adult. Self-assessment, when combined with other treatment elements, reduced disruptive behavior in the trained setting and in near transfer.

A few negative outcomes of self-assessment have been reported. J. Ross, Rolheiser, and Hogaboam-Gray (2002-a) reported a case study of self-assessment in a grade 11 mathematics classroom that resulted in reduced achievement compared to a control class taught by the same teacher (ES=-.35). Interview data suggested that student self-assessments persuaded students that they did not understand core mathematics ideas, even though they were working hard to learn them. The conclusion that many students drew was that they lacked the ability to do advanced level mathematics. Some responded immediately with ego-protecting effort reduction. Others resolved to move out of the advanced mathematics stream in the next school year. However, the credibility of the study was marred by ability differences between the classes, which were resolved only partially by statistical adjustments. Aitchison (1995) assigned grade 7-8 students to a variety of assessment conditions. Students in the self-assessment section scored lower on two instrumental music tasks than students in the assessment by teacher alone condition or in a condition in which students and teacher collaborated on the assessment. However, in the Aitchison study students completed self-assessments on only three occasions and they received no feedback on the accuracy of their judgments.

On balance, the research evidence suggests that self-assessment contributes to higher student achievement and improved behavior. Figure 1 adapted from J. Ross et. (2002-a) provides an explanation for the findings based on social cognition theory (Bandura, 1997).
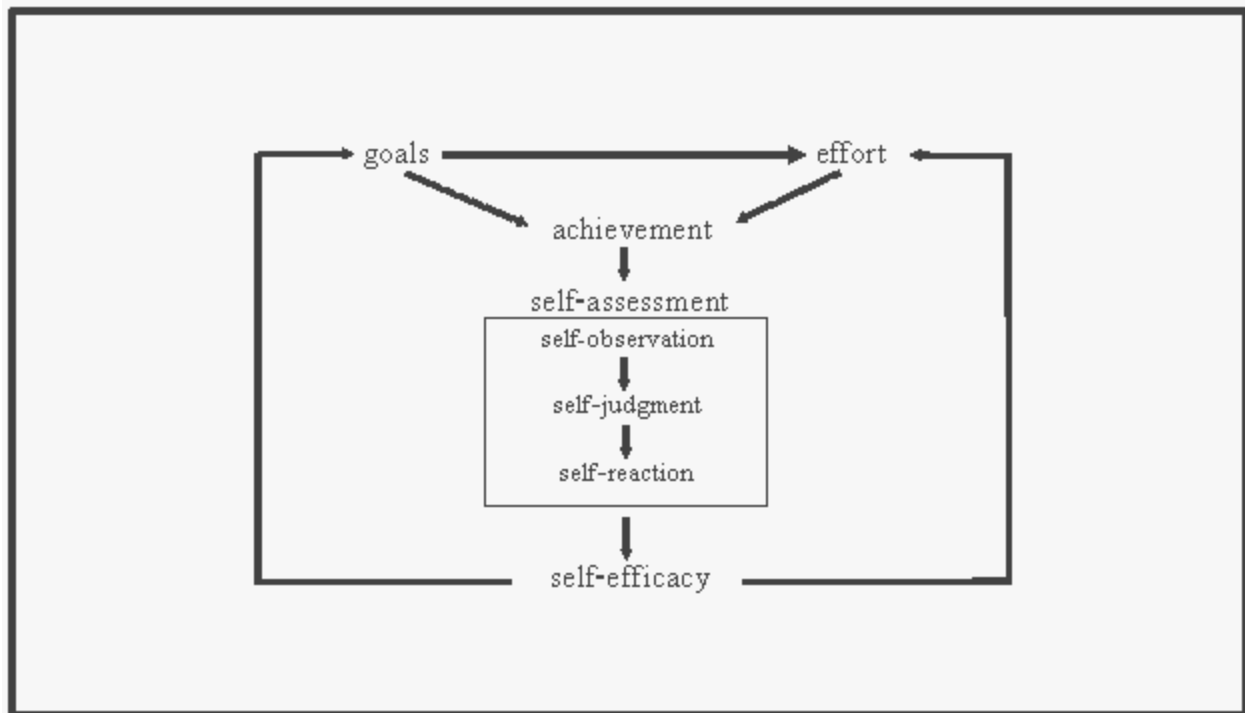
Figure 1: How Self-Assessment Contributes to Learning (adapted from Ross et al., 2002-a)

Self-assessment embodies 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 related 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 achievement that express how satisfied students are with the result of their actions. Training in self-assessment has an impact on students' self-assessments by focusing student attention on particular aspects of their performance (e.g., the dimensions of the co-constructed rubric), by redefining the standards students use to determine whether they were successful (e.g., the levels of the rubric), and by structuring teacher feedback to reinforce positive reactions to the accurate recognition of successful performance. These influences of self-assessment training increase the likelihood that students will interpret their performance as a mastery experience, the most powerful source of self-efficacy information (Bandura, 1997).

Self-assessment contributes to self-efficacy beliefs, i.e., student perceptions of their ability to perform the actions required by similar tasks likely to be encountered in the future. Students who perceive themselves to have been successful on the current task (i.e., who recognize it as a mastery experience) are more likely to believe that they will be successful in the future (Bandura, 1997). Self-assessment training also contributes to self-efficacy through vicarious experience (i.e., classroom discussions of exemplars provide examples of successful experience by students' peers). In addition, the willingness of teachers to share control of assessment constitutes an "inviting message"; i.e., information that the teacher perceives students to be able and responsible, an important source of positive efficacy information (Usher & Pajares, 2005).

Students with greater confidence in their ability to accomplish the target task are more likely to visualize success than failure. 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). The combination of higher goals and increased effort contributes to higher achievement.

Positive self-assessments foster an upward cycle of learning, as demonstrated by the studies that found positive outcomes for self-assessment. But the processes in Figure 1 can generate negative outcomes, as found in J.Ross et al. (2002-a). A stream of negative self-assessments can lead students to select personal goals that are unrealistic, adopt learning strategies which are ineffective, exert low effort and make excuses for performance (Stipek, Recchia, & McClintic, 1992).

Is self-assessment a useful student assessment technique?

Strengths. There is ample evidence that self-assessment contributes to student achievement (Hughes et al., 1985; Schunk, 1996; Sparks, 1991), particularly if teachers provide direct instruction in how to self-assess (e.g., J. Ross et al., 1999; 2002-a; 2005). There is also evidence that self-assessment contributes to improved student behavior (Henry, 1994; Nelson et al., 1995; J. Ross, 1995).

Some data suggest that students prefer self-assessment to assessment by the teacher alone. The reasons given by grade 5-11 students why they preferred self-assessment suggest additional benefits of self-assessment: 1) students said that with self-assessment they had a better understanding of what they were supposed to do because they were involved in setting the criteria for the assessment; 2) students argued the self-assessment was fairer because it enabled them to include important performance dimensions, such as effort, that would not usually be included in their grade; 3) self-assessment enabled them to communicate information about their performance (e.g., their goals and reasoning) that was not otherwise available to their teacher; 4) self-assessment gave them information they could use to improve their work (J. Ross et al., 1998). These changes in perception in the value of assessment through greater involvement in the process might reduce the trend reported by Paris, Lawton, Turner and Roth (1991) in which students become increasingly cynical about the validity and value of assessment as they move through the school system.

Self-assessment encourages students to focus on their attainment of explicit criteria, rather than normative comparisons to other students, (although Blatchford's, 1997 procedure is an exception). For example, when a grade four student in a classroom that used self-assessment extensively was asked what she compared her work to, she reported, "I usually compare it to my own work because not other people's marks are going on my report card…so I need to see if I improved" (J. Ross, Rolheiser, & Hogaboam-Gray, 2002-c, p. 92). The same study found that student conversations about self-assessment were much less focused on marks than their conversations about assessments by the teacher, even though both types of assessment contributed to the final grade.

Relatively little research has been conducted on the benefits of self-assessment for other groups. Teachers might benefit from self-assessment to the extent that making assessment criteria explicit to students might help teachers clarify their intentions and distinguish essential from less important features of student performance. More focused teaching might result. Teacher-student conferences to resolve discrepancies between self- and teacher-assessments might give teachers insights into student thinking, especially student misconceptions that impede further learning. Subsequent instruction might explicitly address deficiencies revealed in the conference. Little has been written about parent reactions to self-assessment. However, the construction of rubrics using language meaningful to students might also make the goals of the curriculum more accessible to parents and the meaning of expected standards more transparent.

Weaknesses. The number one concern of teachers about self-assessment is the fear that sharing control of assessment with students will lower standards and reward students who inflate their assessments. Lack of agreement of self-assessment

with teacher appraisals has many causes. Some are errors of innocence. For example, students conducting retrospective assessments of their work may not recall what they did; they may be unable to accurately assess their production because they have no idea what high performance looks like or they may not understand the assessment criteria or they may lack the deductive skills involved in applying the criteria to their work. But the greatest teacher concern is that "mark sharks" will intentionally inflate their achievement, lying about their effort or misapplying the criteria. Students are also concerned about their ability to self-assess and of the potential for cheating. As one noted, "People could just take advantage of it and just mark all perfect when it's really not their best" (J. Ross et al., 1998).

Even though students prefer self-assessment to teacher appraisal alone, such participation is more work for students. Some describe it as boring (J. Ross et al., 1998) and argue that it is unfair to ask them to do the teacher's job. Teachers express concern about the lack of student commitment to the process, arguing that self-assessment will not work if students do not put the required effort into it.

In some jurisdictions, self-assessments may not be used in the determination of the student's final grade. For example, in one Canadian province "Self Assessment … should **not** be used to inform their report card grade or mark" (Ontario Ministry of Education and Training, p. 463, emphasis in the original). However, self-assessments may be included in a separate learning skills section of the report card. This policy requirement reduces teacher willingness to use self-assessment and may depress student motivation to take the task seriously if they know that it does not count toward their grades.

Although most research on self-assessment focuses on its contribution to academic achievement, some teachers use self-assessment only to measure social skills and to encourage compliance with classroom rules. One explanation might be the previously noted policy impediments to using self-assessment for academic grading. In addition, teachers may find it more challenging to engage students in constructing rubrics for academic performance than for behavioral goals.

Academic rubrics are time consuming to produce. In addition, web-based repositories are not easily accessed; the criteria may too general or too numerous; there may be insufficient delineation between levels and descriptors may be too brief or too long (Dornisch & Sabatini McLoughlin, 2006). However, behavioral rubrics are more familiar to students and easier to use. For example, "comes prepared to class" is a criterion that is easier for teachers to describe and is more easily applied than a criterion like "demonstrates conceptual understanding".

Finally, teachers are concerned about parent reactions to self-assessment. Some parents expect teachers to take sole responsibility for assessment decisions. In a study of conferences involving students, teachers and parents Blake (2000) found that parents expected teachers to "sign off" on student self-assessments, confirming their validity.

Making Self-Assessment More Useful

Teachers who are concerned about the inaccuracy of self-assessment may be partially reassured by the research evidence about the psychometric properties of self-assessment. The concern is likely to remain. Improvement in the utility of self-assessment is most likely to come from attention to four dimensions in training students how to assess their work.

First, the process for defining the criteria that students use to assess their work will improve the reliability and validity of assessment if the rubric uses language intelligible to students, addresses competencies that are familiar to students, and includes performance features they perceive to be important. Rolheiser (1986) suggested several strategies for engaging students in the construction of simple rubrics. A key message in Rolheiser's manual is that teachers should not surrender control of assessment criteria but enact a process in which students develop a deeper understanding of key expectations mandated by governing curriculum guidelines. Offering to expand the rubric to include additional "kid-criteria" contributes to student commitment. In addition to focusing student attention on specific aspects of a domain, the construction of a rubric also provides students with

a language for talking about their learning. In some instances, a process of progressive revelation of the rubric may be appropriate, if students lack sufficient experience in the domain to be able to identify dimensions of mastery.

Second, teaching students how to apply the criteria also contributes to the credibility of the assessment and student understanding of the rubric. Among the more powerful strategies are teacher explanations of each criterion, teacher modeling of criteria application, and student practice in applying the rubric to examples of student work (including their own). Within-lesson comments that link instructional episodes and student tasks to assessment criteria reinforce student understanding of the criteria.

Third, giving students feedback on their self-assessments is a process of triangulating student self-assessments with teacher appraisals and peer assessments of the same work using the same criteria. Conferencing with individuals and groups to resolve discrepancies can heighten attention to evidence, the antidote to lying and self-delusion. A key issue is to help students move from holistic to analytic scoring of their work. For example, student self-assessments are frequently driven by their perception of the effort expended on the assignment, an important criterion but it should not swamp attention to other dimensions of performance.

Fourth, students need help in using self-assessment data to improve performance. Student sophistication in processing data improves with age. For example, J. Ross et al. (2002-c) found that when discussing assessments with parents and peers, grade 6 students were more likely to focus on evidence of achievement and how to improve performance, whereas grade 2-4 students focused exclusively on the overall grade. In addition, older students were more likely than younger to compare current to past achievement on similar tasks. Teachers can provide simple recording forms for tracking performance over time to compensate for memory loss. Teachers can provide games, conferences, and menus of examples to support goal setting. Goals are more likely to improve student achievement if they are set by students

themselves, are specific, attainable with reasonable amounts of effort, focus on near as opposed to distant ends, and link immediate plans to longer term aspirations. Recording goals in a contract increases accountability. Teachers can also address student beliefs that contribute to higher goal setting, such as attributions for success and failure and seeing ability as something that can improve rather than as a fixed entity.

Teachers can further support self-assessment by creating a climate in which students can publicly self-assess. Strategies for creating trust in the classroom are readily available (e.g., Johnson, Johnson, & Holubec, 1990). The usefulness of self-assessment is likely to be enhanced by strategies that shift students toward learning goals (i.e., students approach classroom tasks in order to understand key ideas) and away from performance goals (i.e., students approach classroom tasks in order to demonstrate they are smarter than their peers).

Conclusions

There is sufficient information from research on self-assessment to answer the questions posed at the outset of this article with reasonable confidence. The psychometric properties of self-assessment suggest that it is a reliable assessment technique producing consistent results across items, tasks and contexts and over short time periods. Teachers can strengthen reliability through such strategies as engaging students in rubric construction. Evidence of the validity of self-assessment is mixed. Self-assessments are typically higher than teacher assessments, although the size of the discrepancy can be reduced through student training and other teacher actions. In addition, differences between self- and teacher-assessment can lead to productive teacher-student conversations about student learning needs. There is persuasive evidence, across several grades and subjects, that self assessment contributes to student learning and that the effects grow larger with direct instruction on self assessment procedures. The central finding of this review is that the strengths of self-assessment can be enhanced through specific student training and each of the weaknesses of the approach (including inflation of grades) can be reduced through teacher action.

Teachers can learn more about how to teach students the skills of self-assessment by consulting manuals (e.g., Rolheiser, 1996) and through in-service activities. Little research has been conducted on strategies for training teachers in this domain, with the exception of J. Ross et al. (1998). This study compared two in-service options: a skills training approach (researchers presented strategies for teaching self-assessment) and an action research model (teacher developers of a self-assessment program served as mentors for in-service participants to develop their own programs). The study found that the action research version of the in-service had a more positive effect than the other condition on student attitudes toward assessment, in part because the learner control of the action research approach to in-service was congruent with sharing control with students in self-assessment.

The research reviewed in this article provides teachers with evidence that self-assessment, when properly implemented, produces valid and reliable information about student achievement. The optimal use of this information is for formative purposes, providing credible data on student cognitions about their achievement that is not otherwise available to teachers. Teachers who make a serious commitment to learning about self-assessment and teaching these techniques to their students can plausibly anticipate enhanced student motivation, confidence, and achievement.

Notes:

[2] Not all supporters of assessment reform subscribe to consequential validity. Popham (1997) argued that the consequences of assessment should be a major concern of right minded people but cluttering the concept of validity with issues of the effects of test use sows confusion. Popham argued that test developers should be concerned with consequences, intended and unintended, but a test can be valid or invalid regardless of its consequences. Others, like Messick (1995) argue that adverse consequences undermine validity only if the problem is the result of a poor fit of the test with what it purports to measure.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*, 587-605.

Aitchison, R. E. (1995). *The effects of self-evaluation techniques on the musical performance, self-evaluation accuracy, motivation and self-esteem of middle school instrumental music students.* Unpublished doctoral dissertation, University of Iowa.

Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21-34.

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 annual meeting of the American Educational Research Association, New Orleans.

Aschbacher, P. R. (1991). Performance assessment: State, activity, interest, and concerns. *Applied Measurement in Education, 4*, 275-288.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Bergee, M. J. (1997). Relationships among faculty, peer, and *self*-evaluations of applied performances . *Journal of Research in Music Education, 45*(4), 601-612.

Blake, M. (2000). Developing a curriculum linking cooperative learning and student-directed parent-teacher conferences. *Schools in the Middle, 9*(7), 40-44.

Blatchford, P. (1997). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology, 17*(3), 345-360.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education, 18*, 529-549.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience and school*. Washington, DC: National Academy Press.

Butler, R. (1990). The effects of mastery and competitive conditions on self-assessment at different ages. *Child Development, 61*, 201-210.

Cassady, J. C. (2001) *Self-reported GPA and SAT: A methodological note. Practical Assessment, Research & Evaluation, 7(12).* [Web Page]. URL Retrieved November 14, 2006 from http://PAREonline.net/getvn.asp?v=7&n=12.

Dornisch, M. M., & Sabatini McLoughlin, A. (2006). Limitations of web-based rubric resources: Addressing the challenges . *Practical Assessment, Research & Evaluation, 11*(3).

Fitzgerald, J. T., Gruppen, L. D., & White, C. B. (2000). The influence of task formats on the accuracy of medical students' self-assessments . *Academic Medicine, 75*(7), 737-741.

Fontana, D., & Fernandes, M. (1994). Improvements in math performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology, 64*, 407-417.

Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology, 41*, 581-592.

Gregory, K., Cameron, C., & Davies, A. (2000). *Self-assessment and goal-setting*. Merville BC: Connections Publishing.

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

Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.

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

Johnson, D. W. J. R., & Holubec, E. (1990). *Circles of learning: Cooperation in the classroom*. Edina, MN: Interaction Book Company.

Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., & Eisenberg, S. N. (2004). School-age children's self-assessment of oral narrative production . *Communication Disorders Quarterly, 26*(1), 37-48.

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education, 2*(2), 145-163.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis. *Review of Educational Research, 75*(1), 63-82.

Longhurst, N., & Norton, L. S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation, 23*(4), 319-330.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education, 10*(2), 209-220.

McEnery, J. M., & Blanchard, P. N. (1999). Validity of multiple ratings of business student performance in a management simulation. *Human Resource Development Quarterly., 10*(2), 155-172.

McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction. 3rd edition*. Toronto: Pearson Allyn Brown.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment.

Educational Measurement: Issues and Practice, 14(4), 5-8.

Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice, 17*(2), 6-12.

Nelson, J. R., Smith, D. J., & Colvin, G. (1995). The effects of a peer-mediated self-evaluation procedure on the recess behavior of students with behavioral problems. *Remedial and Special Education, 16*(2), 117-126.

Newman, F. (1997). Authentic assessment in social studies: Standards and examples. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, adjustment, and achievement*. San Diego: Academic Press.

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools . *Practical Assessment, Research & Evaluation, 10* (17).

Ontario Ministry of Education and Training. (2005). *Ontario provincial elementary assessment and evaluation: Effective elementary assessment and evaluation classroom practices.* Toronto: author.

Pakaslahti, L., & Keltikangas-Järvinen, L. (2000). Comparison of peer, teacher and self-assessments on adolescent direct and indirect aggression. *Educational Psychology, 20*(2), 177-190.

Pappas, G., Lederman, E., & Broadbent, B. (2001). Monitoring Student Performance in Online Courses: New Game-New Rules. *Journal of Distance Education, 16*(2).

Paris, S., Lawton, T., Turner, J., & Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*(5), 12-20, 40.

Phillips, D. A., & Zimmerman, M. (1990). The development course of perceived competence and incompetence among competent children. In R. J. Sternberg & J. J. Kolligian (Eds.), *Competence considered* (pp. 41-66). London: Tale University Press.

Popham, W. J. (1997). Consequential validity: Right concern--wrong concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13.

Rolheiser, C. (Ed.). (1996). *Self-evaluation...Helping kids get better at it: A teacher's resource book.* Toronto: OISE/UT.

Ross, J. A. (1995). 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. A., Hogaboam-Gray, A., & Rolheiser, C. (2002a). Self-Evaluation in grade 11 mathematics: Effects on achievement and student beliefs about ability. In D. McDougall (Ed.), *OISE Papers on Mathematical Education*. Toronto: University of Toronto.

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002b). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment, 8*(1), 43-58.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effect of self-evaluation on narrative writing. *Assessing Writing, 6*(1), 107-132.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002c). Influences on student cognitions about evaluation. *Assessment in Education, 9*(1), 81-95.

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

Ross, J.A. & Starling, M. (2005, April). *Achievement and self-efficacy effects of self-evaluation training in a computer-supported learning environment.* Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15*(1), 1-20.

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill

learning. *American Educational Research Journal, 33*(2), 359-382.

Sparks, G. E. (1991). *The effect of self-evaluation on musical achievement, attentiveness and attitudes of elementary school instrument students.* Unpublished doctoral dissertation, Louisiana State University and Agricultural and Mechanical College.

Stipek, D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development, 57*, 1-84.

Sundström, A. (*Self-assessment of knowledge and abilities: A literature study* [Web Page]. URL http://www.umu.se/edmeas/publikationer/pdf/EM541.pdf [2006, November 14].

Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education, 45*(2), 187-202.

Talento-Miller, E., & Peyton, J. (2006) *Moderators of the accuracy of self-report Grade Point Average. General Management Admission Council Research Reports • RR-06-10* [Web Page]. URL http://www.gmac.com/NR/rdonlyres/9B6C60A5-F9D6-4E69-91CB-22F949E7CA59/0/RR0610_SRUGPA.pdf [2006, November 14].

Usher, E. L., & Pajares, F. (2005, April). *Inviting confidence in school: Sources and effects of academic and self-regulatory efficacy beliefs.* Paper presented at the meeting of the American Educational Research Association, Montreal.

Wiggins, G. (1993). *Assessing student performance: Explore the purpose and limits of testing.* San Francisco, CA: Jossey-Bass.

Wiggins, G. (1998). *Educative Assessment: Designing assessments to inform and improve student performance.* San Francisco: Jossey-Bass.

.

Citation

Author

Dr. John A. Ross,
Professor and Centre Head,
PO Box 719, 1994 Fisher Dr.
Peterborough, ON K9J 7A1
Canada