21

Student Self-Assessment[1]

Gavin T. L. Brown

Lois R. Harris

The only way any of us can improve—as Coach Graham taught me—is if we develop a real ability to assess ourselves. If we can't accurately do that, how can we tell if we're getting better or worse? (Pausch & Zaslow, 2008, p. 112)

During the past two decades, student self-assessment has been strongly endorsed as an important aspect of formative assessment through the global assessment for learning (AFL) movement. Student self-assessment is not new, with Brookhart (2009) noting that even in the 1930s and 1940s there were numerous authors endorsing the use of student self-evaluation. However, self-assessment is seldom implemented in many classrooms. Hunter, Mayenga, and Gambell (2006) found that 23% of the 4,148 Canadian secondary teachers sampled reported never using self-assessment, with 58% reporting minimal self-assessment use. Only half of 346 surveyed upper secondary students in Finland reported participating in self-assessment (Lasonen, 1995). This limited implementation likely relates to the tensions teachers report between the use of student-led assessment practices and the externally and teacher-controlled summative results generally reported to stakeholders (Harris & Brown, 2010; Volante & Beckett, 2011).

There is general consensus that self-assessment is positive and leads to benefits for students. Perhaps the most powerful promise of self-assessment is that it can raise student academic performance by teaching pupils self-regulatory processes, allowing them to compare their own work with socially defined goals and revise accordingly (Andrade, 2010; Black & Wiliam, 1998; Butler & Winne, 1995; Hattie & Timperley, 2007; Ramdass & Zimmerman, 2008). The logic is that, like self-regulation, self-evaluation of the quality attributes of one's own work draws on metacognitive competencies (e.g., self-observation, self-judgment, self-reaction, task analysis, self-motivation, and self-control) (Zimmerman,

[1]Acknowledgment: We wish to acknowledge the financial support of the Assessment Research Centre grant #C7-R8000-0, Hong Kong Institute of Education that permitted the employment of Ka Ki Kitty Cheng as a research assistant. The first draft of this chapter was written while the first author was employed at the Hong Kong Institute of Education.

2002). Additionally, there is evidence that students can improve their self-regulation skills through self-assessment (i.e., set targets, evaluate progress relative to target criteria, and improve the quality of their learning outcomes) (Andrade, Du, & Mycek, 2010; Andrade, Du, & Wang, 2008; Brookhart, Andolina, Zuza, & Furman, 2004). Furthermore, self-assessment is associated with improved motivation, engagement, and efficacy (Griffiths & Davies, 1993; Klenowski, 1995; Munns & Woodward, 2006; Schunk, 1996), reducing dependence on the teacher (Sadler, 1989). It is also seen as a potential way for teachers to reduce their own assessment workload, making students more responsible for tracking their progress and feedback provision (Sadler & Good, 2006; Towler & Broadfoot, 1992).

This chapter reviews relevant empirical studies concerning the use of student self-assessment in the compulsory school sector (K–12) to help establish which claims about self-assessment are empirically supported. Previous reviews of this topic have focused primarily on higher education students (e.g., Boud & Falchikov, 1989; Dochy, Segers, & Sluijsmans, 1999; Falchikov & Boud, 1989; Mabe & West, 1982). This chapter contributes to our understanding of self-assessment in public schooling.

Defining Self-Assessment

Many terms have been used to describe the process of students assessing and providing feedback on their own work, including self-assessment, self-evaluation, self-reflection, self-monitoring, and more generally, reflection. Since self-assessment requires evaluative consideration of one's own work, the processes of self-grading, self-testing, and self-rating can also potentially be forms of self-assessment. Both the *Thesaurus of ERIC Descriptors* (Educational Research Information Center [U.S.], 2001) and the *Thesaurus of Psychological Index Terms* (Tuleya, 2007) treat self-assessment as a synonym for self-appraisal, and both are classified under the subject heading *self-evaluation (individuals)* or *self-evaluation,* respectively. The *Thesaurus of ERIC Descriptors* defines self-evaluation as "individuals' assessment of themselves" (Educational Research Information Center [U.S.], 2001).

When examining the word *assessment*, the Joint Committee on Standards for Educational Evaluation (2003) defines it as a "process of collecting information about a student to aid in decision making about the student's progress and development" (p. 5). Accepting this definition of assessment, then logically, self-assessment must involve students collecting data to evaluate their own progress, consistent with Klenowski's (1995) statement that self-evaluation requires students "to evaluate and monitor their own performance in relation to identified criteria or standards" (p. 146). Hence, within a compulsory school setting and when serving academic purposes, we take a global and generic approach that self-assessment is a descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities.

Self-assessment can be operationalized in many ways, ranging from a careful consideration of the quality of one's own work guided by a rubric or feedback from the teacher, to scoring one's own work, to practices like predicting one's likely score on an impending task or test. What distinguishes these actions from other assessment practices is that they are carried out by the student (Brooks, 2002), though the degree of autonomy from peers, teachers, or parents will vary in practice. Unlike Boud and Falchikov (1989), who privileged techniques that require an evaluative, criterion-based judgment, we have not excluded self-marking or self-rating techniques. Instead of restricting self-assessment to solely the act of evaluating the quality of work against socially agreed criteria, we include self-assessment acts that involve estimating quantitative aspects of work (e.g., amount, speed, score, or place on a hierarchy/progression). This gives us a broad scope to establish whether there are different effects depending on the type of self-assessment carried out. Thus, self-assessment takes place when students impute or infer that their work or their ability to do that work has some sort of quality characteristics, and this self-assessment may, in its most simple form, be a quantity estimate (i.e., How many task requirements have I satisfied?) or a quality estimate (i.e., How well have I done?).

In taking this broad stance toward self-assessment, we are aware that not all scholars share our perspective. Some classroom assessment

(CA) researchers (e.g., Andrade, 2010) make a robust distinction between assessment and evaluation in which the latter is considered to refer to grading, testing, or marking (hence, summative) rather than the more formative, improvement-oriented emphasis implied by assessment. Other authors (e.g., Clarke, Timperley, & Hattie, 2003) have prioritized a child-centered pedagogical process in which self-assessment focuses the student on processes that lead to improved outcomes without focusing on an evaluative dimension.

Distinguishing between assessment and evaluation has become commonplace in the AFL community, partly as a consequence of Sadler's (1989) assertion that formative and summative evaluations were qualitatively different forms of assessment. This stands in contrast to Scriven's (1967) definition, which focuses on the timing of the interpretations and uses of assessment rather than its form. While agreeing that formative improvement is the fundamental purpose for using any type of assessment (Popham, 2000), it is our position that there is little merit in creating a dichotomy between assessment and evaluation, because all assessments, including formative ones, describe and evaluate the merit, worth, or quality of student work (Hattie & Brown, 2010). Consequently, studies that involve compulsory school students making judgments about their own work or academic ability using a diverse range of assessment methods (e.g., tests, graded assignments, essays, performance tasks, or rubric-guided judgments) have been included in this review.

However, there are limits to what we would consider to be self-assessment. As Kasanen and Räty (2002) pointed out, within an academic context, self-assessment is not about the process of knowing oneself better, as in the notion that an unexamined life is not worth living; rather, it is about judging, evaluating, and considering one's own academic work or abilities. Hence, in this chapter we do not include studies primarily concerned with how children evaluate their self-concept or self-worth (e.g., Burnett, 1996; Byrne & Bazana, 1996; Marsh, 1988; Williams, 1996). The focus in this chapter is on self-assessment of schoolwork, rather than on personal well-being, because a major focus of schools is to help students learn new skills and knowledge and develop their understanding of school curriculum materials.

Self-Assessment Techniques

Methods of Self-Assessment

Research studies tend to emphasize self-assessment methods that focus directly on obtaining from students an estimate or description of how well they believe they will do or have done on a specific test or task. In general, it seems that self-assessment practices can be grouped into three major types: (1) self-ratings, (2) self-estimates of performance, and (3) criteria- or rubric-based assessments. Self-assessment practices may also encourage students to include comments or advice from the student to him or herself about how to improve.

Self-rating requires students to judge quality or quantity aspects of their work using a rating system. In the classroom, checklists that remind students of important task characteristics or task processes are commonplace (e.g., Clarke et al., 2003). Clarke (2005) has also created self-rating prompts that are more evaluative and task-oriented—that is, "a) I am pleased with my work because I . . . ; b) Two improvements I have made are . . . ; c) I would grade myself A B C D E because I . . . ; and d) Next time I need to focus on . . ." (p. 113). In Clarke's (2005) approach to self-rating, students are providing not only a rating of the quality of their work but are also expected to give feedback comments for improvement, mimicking the formative feedback teachers might provide. Another commonplace rating technique is the use of *traffic lights*, where students show the teacher a red, yellow, or green circle to indicate readiness to proceed or quality of understanding, with red signifying difficulty and green meaning comprehension (Black & Harrison, 2001; Clarke, 2005).

Self-marking or grading of one's own work can also be done using either a marking guide for objectively answered questions or a rubric or model answer (Todd, 2002). While some simple self-rating practices like self-marking have sometimes been shown to be trustworthy (Wall, 1982), some teacher educators (e.g., Brooks, 2002) consider that simple mechanical marking (e.g., right versus wrong) is unlikely to be effective in improving learning since high levels of cognitive engagement are absent. Alternatively, self-assessment may involve students estimating their level of performance or ability relative to a test or a task they are about to take, have just

taken, or recall having taken some time previously (e.g., How well have I done on this test?). Some of these self-assessments are more global and may also require students to mentally estimate how well they performed on a test in the form of a test mark or score, a rank order position, or a grade.

Lastly, and perhaps most classically associated with AFL, is the practice of using a rubric to ascertain the quality characteristics of the individual's written or performed work. Rubrics may or may not have score indicators (e.g., A, Level 3, or excellence) but always arrange quality indicators in incremental progressions that students use to best fit the various aspects of their work. Rubrics are especially common when students are assessing writing or judging portfolios or collections of work (Andrade & Valtcheva, 2009). These three types of self-assessment show that self-assessments can be global (e.g., How good is my writing?) or anchored to a specific task (e.g., How well did I do on question 3?) as all such tasks require reflection on the quality of the student's work.

Accuracy in Self-Assessment

The role of accuracy of self-assessment is contentious. Brooks (2002) has argued that reliability matters for external assessments, not for formative classroom purposes. Others have argued that grading one's own work (Lipnevich & Smith, 2008) and being required to conform to a teacher's assessment of the student's work (Paulhus, 1991) have negative effects on students' judgments and undermine the constructive processes of self-regulation. Despite evidence students may be motivated to inflate their grades (Harris & Brown, 2010), there is a minority position (e.g., Chang & Tseng, 2011) that advocates using student self-assessments when determining final results so students feel their judgments are valued.

We believe that accuracy is an important facet in determining the validity of any assessment (Messick, 1989) since accurate self-evaluation is a key component within models of self-regulation of learning (Schunk, 1996; Zimmerman, 1998). Thus, from both psychometric and learning theory perspectives, the accuracy of self-assessment is critical, as suggested by the quote at the beginning of the chapter from Pausch, a computer science professor who used self-assessment principles in his teaching. If self-assessment processes lead students to conclude wrongly that they are good or weak in some domain and they base personal decisions on such false interpretations, harm could be done—even in classroom settings (e.g., task avoidance, not enrolling in future subjects) (Ramdass & Zimmerman, 2008).

Consistent with reliability theory (Haertel, 2006), we consider that all self-assessments, no matter how privileged the self is in terms of knowing what the self has done, are imperfect indicators of competence. Indeed, Dunning, Heath, and Suls (2004) identified many reasons self-assessments can be flawed. These include a tendency for humans (1) to be unrealistically optimistic about their own abilities (e.g., "I can finish this in just one week"), (2) to believe that they are above average (e.g., no one admits to being a poor driver, lover, or friend), (3) to neglect crucial information (e.g., ignore key performance indicators that should be used to evaluate their work), and (4) to have deficits in their information (e.g., simply do not know what to look for in determining the quality of their work). Furthermore, lack of competence in a domain (as would be expected in a low progress learner or beginner) has a dual handicapping effect; such people are not very good in the domain and, at the same time, are not aware that they are not good in the domain (Dunning et al., 2004). Additionally, pressure to enhance one's own self-worth may result in overestimation of ability (Saavedra & Kwun, 1993) and inaccurate self-reporting of grades or test scores (Kuncel, Credé, & Thomas, 2005). Students have also been found to take their own effort, which ought to be independent of quality, into account when evaluating their work (Ross, Rolheiser, & Hogaboam-Gray, 1998b). In much simpler terms, as Dr. Gregory House of *House* puts it, "Everybody lies" (Ruff & Barris, 2009, p. 84).

Another pressure on accurate self-assessment is that much of what makes one competent in many domains is relatively ill defined. Consider the great difficulty teachers have in scoring student work against standards or rubrics (Brown, 2009), often providing inaccurate or inconsistent judgments of student work (Topping, 2003). Thus, without putting any responsibility or blame on students, there are many good reasons to expect that their self-assessments of their own work products or performances will be reasonably

flawed or inaccurate. Hence, while self-assessment has considerable promise for helping students improve their learning within compulsory school settings, it is not without potential problems and limitations.

Literature Review Method

Selection of Studies

In this chapter, we examined studies of student self-assessment carried out in the compulsory school sector to discern which claims can be empirically supported. Since self-assessment is subsumed by the term *self-evaluation,* that subject heading was initially used to query the Education Resources Information Center (ERIC) and PsycINFO databases. The search initially identified 348 potentially relevant sources. Through a check of abstracts and titles, studies were excluded from this sample when they were the following:

- Not readily available from the authors or the Internet
- In languages other than English
- Conducted outside the K–12 sector (e.g., higher education)
- Related primarily to student self-concept
- Conceptual, not empirical, analyses of self-assessment
- Related specifically to special education (important but outside the scope of this review)

Further searches carried out in these databases, using the key words *self-assessment* and *schools,* resulted in the collection of 11 additional relevant studies. Papers cited in existing reviews (e.g., Andrade & Valtcheva, 2009; Black & Wiliam, 1998; Ross, 2006) were also collected. The current paper provides a synthesis of 84 empirical studies on student self-evaluation in compulsory education.

Research Questions

Drawing on the claims and quality issues raised in the previous section, we reviewed the empirical literature in light of the following questions:

1. What is the relationship between self-assessment and student academic achievement?
2. What is the relationship between self-assessment and self-regulation (including motivation or engagement)?
3. How do students perceive and experience self-assessment?
4. What are the relationships between self-assessment accuracy and student age and ability?
5. What are the relationships between task features, method of self-assessment, and self-assessment accuracy?

Analysis

Studies were read and assigned to thematic categories arising from the research questions of the paper: (1) relationship to academic performance or achievement (achievement); (2) relationship to self-regulating processes (self-regulation); (3) student perspectives (student perspectives); and (4) accuracy concerns relative to student age, student experience, student proficiency, task characteristics, or means of self-assessment (accuracy). Both authors agreed on the classifications of each study.

Where sufficient data were provided, Cohen's (1992) *d* effect sizes (i.e., a standardized measure of difference as a proportion of standard deviation) were computed using an Excel macro developed by Wilson (2001). These standardized effect sizes allow the overall impact of the self-assessment practices described in the studies to be compared within this data set (e.g., average effects on different sample populations) and against effect sizes that have been computed for other educational practices. Within education, the average of all interventions reviewed in a large-scale synthesis of meta-analyses has been estimated to be $d = 0.40$ and values ≥ 0.60 are considered large (Hattie, 2009).

Empirical Evaluation of Self-Assessment in Education

The main design, demographic, and thematic content of every reviewed study has been summarized in Table 21.1.

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
1	Alsaker (1989)	Survey	Norway	Middle school (G6–9)	2,309	Examined relationships between self-esteem, perceived academic competence, school importance, and achievement	Accuracy
2	Andrade & Boulay (2003)	Quasi-experiment	United States	Middle school (G7–8)	397	Compared writing results of an experimental group (rubric-guided self-assessment training) with a control group (rubric access only)	Achievement
3	Andrade, Du, & Mycek (2010)	Quasi-experiment	United States	Elementary & middle school	162	Investigated the relationship between writing scores and modeling, generating criteria and self-assessing using a rubric	Achievement
4	Andrade, Du, & Wang (2008)	Quasi-experiment	United States	Elementary school (G3 & 4)	116	Evaluated the effects of modeling, generating criteria, and self-assessing using a rubric on writing performance	Achievement
5	Andrade, Wang, Du, & Akawi (2009)	Quasi-experiment	United States	Elementary & middle school (G3–7)	268	Investigated rubric and self-assessment usage in writing, examining effects on self-efficacy with focus on gender	Self-efficacy
6	Ash (1980)	Survey	United States	High school	156	Examined the usefulness of typing ability self-assessment	Accuracy
7	Barling (1980)	Experiment	South Africa	Elementary school (G3–6)	138	Compared effects of five different treatments (control, control plus feedback, self-monitoring, self-determined performance standards and self-reinforcement, or self-instruction) on academic performance	Achievement
8	Barnett & Hixon (1997)	Quasi-experiment	United States	Elementary school (G2, 4, 6)	62	Investigated how grade level and subject related to students' ability to predict test scores	Accuracy
9	Blatchford (1997a)	Longitudinal survey	United Kingdom	Elementary—high school (ages 7–16)	108	Explored the accuracy and stability over time of students' academic self-assessments	Accuracy
10	Blatchford (1997b)	Longitudinal survey	United Kingdom	Elementary & high school	108	Examined the effects of sex and ethnicity on academic attainment self-assessments	Accuracy

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
11	Bradshaw (2001)	Quasi-experiment	United States	Elementary school (G3–5)	87	Compared student ratings with actual performance	Accuracy
12	Brookhart, Andolina, Zuza, & Furman (2004)	Action research	United States	Elementary school (G3)	41	Investigated self-assessment of strategy use for memorizing and recalling mathematics facts, learning reflection, and using metacognitive skills	Self-regulation; Student perspectives
13	Brown, Irving, Peterson, & Hirschfeld (2009)	Survey	New Zealand	High school (G9–10)	705	Mapped relationships between student definitions and conceptions of assessment	Student perspectives
14	Brown, Peterson, & Irving (2009)	Survey	New Zealand	High school (G9–10)	624	Analyzed relationships between student definition of assessment responses and mathematics achievement	Student perspectives
15	Butler (1990)	Experiment	Israel	Elementary school (GK, 2, & 5)	80	Examined the effects of age and conditions (mastery and competition) on self-assessment accuracy	Accuracy
16	Butler & Lee (2006)	Survey	Korea	Elementary school	151	Investigated the validity of students' self-assessments of oral English performance	Accuracy
17	Chang & Tseng (2011)	Experiment	Taiwan	High school (G8)	60	Examined the effects of students' use of a Web-based portfolio system that incorporated self- and peer assessment tasks	Accuracy
18	Claes & Salame (1975)	Quasi-experiment	Canada	High school (G9, 10)	65	Compared the accuracy of students' self-evaluations of performance on tasks to the students' overall achievement.	Accuracy
19	Connell & Ilardi (1987)	Survey	United States	Elementary school (G4–6)	121	Investigated variables in children's self-ratings of academic competence	Accuracy
20	Cowie (2009)	Multimethod	New Zealand	Elementary, & middle school	22 classes	Investigated student perceptions of experienced formative assessment practices	Student perspectives

(Continued)

Table 21.1 (Continued)

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
21	Daiute & Kruidenier (1985)	Experiment	United States	Middle & high school (G7, 9)	57	Evaluated the effect of using computer self-questioning prompts on writing achievement	Achievement
22	Eccles, Wigfield, Harold, & Blumenfeld (1993)	Survey	United States	Elementary school (G1, 2, 4)	865	Examined students' perceived levels of competence in multiple domains (e.g., sport, music, reading, and math)	Accuracy
23	Elder (2010)	Interview	United States	Elementary school (G1 & G4, 5)	37	Investigated how students self-assess schoolwork	Accuracy
24	Fernandes & Fontana (1996)	Quasi-experiment	Portugal	Elementary school (G3, 4)	354	Examined student self-efficacy beliefs and academic achievement after using self-assessment techniques	Self-regulation
25	Frey & Ruble (1987)	Interview & observation	United States	Elementary school (GK–4)	83	Explored relationships between student self-estimates of ability and their age and sex	Accuracy
26	Gao (2009)	Case study	Hong Kong	High school (G10, 11)	1 class	Examined student perceptions of school-based assessment practices, including self-assessment	Student perspectives
27	Glaser, Kessler, Palm, & Brunstein (2010)	Experiment	Germany	Elementary school (G4)	105	Investigated the effects of self-regulation and evaluation training on writing achievement and efficacy	Self-regulation Achievement
28	Griffiths & Davies (1993)	Action research	UK	Elementary school (G5, 6)	1 class	Explored student self-assessment reflections about the learning process	Self-regulation
29	Harris & Brown (2010)	Interview	New Zealand	Elementary, middle, & high school (G5–7, 10)	40	Studied student and teacher perspectives and experiences of self-assessment practices	Student perspectives

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
30	Harris, Harnett, & Brown (2009)	Focus group interviews	New Zealand	Elementary, middle, & high school (G5/6, 7, 10)	46	Examined student perspectives of experienced classroom assessment (CA) practices	Student perspectives
31	Harward, Allred, & Sudweeks (1994)	Experiment	United States	Elementary (G4)	209	Assessed the effects of four self-corrected test methods on spelling achievement	Achievement
32	Hewitt (2001)	Experiment	United States	Middle, & high school (G7–9)	82	Examined the effects of modeling, self-evaluation, and self-listening on junior high school instrumentalists' music performance	Achievement
33	Hewitt (2005)	Survey	United States	Middle & high school	143	Compared the accuracy of student musical performance self-assessments to expert evaluations	Accuracy
34	Higgins, Harris, & Kuehn (1994)	Survey	United States	Elementary school (G1, 2)	46	Compared student self-ratings of their projects on 5-point scale with teacher ratings	Accuracy
35	Hughes, Sullivan, & Mosley (1985)	Experiment	United States	Elementary school (G5)	250	Examined the effects of self versus teacher assessment and task difficulty on student motivation	Self-regulation
36	Ikeguchi (1996)	Survey	Japan	High school	34	Compared students' self-assessments of language skills with teacher assessments and test scores	Accuracy
37	Johnson & Winterbottom (2011)	Interviews & observation	United Kingdom	High school	28	Examined the motivation and perspectives of students in a girls only biology class implementing self- and peer assessment	Self-regulation Student perspectives
38	Jones, Trap, & Cooper (1977)	Quasi-experiment	United States	Elementary school (G1)	22	Compared student self-rating and recording of handwriting performance with adult ratings	Accuracy

(Continued)

Table 21.1 (Continued)

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
39	Kaderavek, Gillam, Ukrainetz, Justice, & Eisenberg (2004)	Survey	United States	Elementary school (ages 5–12)	401	Examined student self-rating of their narratives relative to actual performance	Accuracy
40	Kasanen & Räty (2002)	Multimethod	Finland	Elementary school (G1)	21	Studied the classroom implementation of self-assessment	Student perspectives
41	Kasanen, Räty, & Eklund (2009)	Interview	Finland	Elementary school (G3, 6)	58	Investigated pupils' evaluations of their academic abilities and potential for improvement in different school subjects	Accuracy
42	Keil, McClintock, Kramer, & Platow, (1990)	Quasi-experiment	United States	Elementary & middle School (G2, 4, 6, 8)	480	Examined the effects of information about children's own outcomes and those of a peer on their self-evaluations	Accuracy
43	Klenowski (1995)	Case Study	Australia & United Kingdom	High school	Not given	Investigated how students self-evaluated work and how teachers implemented these practices	Self-regulation
44	Koivula, Hassmén, & Hunt (2001)	Experiment	Sweden	High school (final year)	550	Compared academic results of students, who completed confidence ratings of answers to standardized test items with normal test takers	Accuracy; Achievement
45	Kwok & Lai (1993)	Survey	Canada & Hong Kong	Elementary school	253	Examined student self-evaluations of academic competence relative to their actual performance in mathematics	Accuracy; Student perspectives
46	Lasonen (1995)	Survey	Finland	High school	346	Explored student experiences of self-assessment and the criteria they use to make judgments	Student perspectives

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
47	Laveault & Miles (2002)	Survey	Canada	Elementary & middle school (G5–8)	770	Investigated the relationship between writing proficiency and accurate self-assessment using a rubric	Accuracy
48	LaVoie & Hodapp (1987)	Survey	United States	Elementary school (G4–6)	311	Compared children's perceptions of standardized test performance with actual performance	Student perspectives; Accuracy
49	Lee & Gavine (2003)	Experiment	Scotland	Middle school (G7)	56	Evaluated the effectiveness of an intervention involving students setting goals and participating in self-evaluation	Accuracy
50	Luyten & Dolkar (2010)	Survey	Bhutan	High school (G10)	365	Compared student and teacher ratings to examination scores	Accuracy
51	McDevitt et al. (2008)	Survey	United States	Middle school (G6–8)	90	Investigated the use of self-assessment and goal setting in reading	Student perspectives; Self-regulation
52	McDonald (2002)	Survey	Barbados	High school	570	Explored student definitions of self-assessment	Student perspectives
53	McDonald (2009)	Mixed methods	Caribbean	High school	515	Compared qualitative data about male student and teacher experiences of self-assessment training with examination results	Self-regulation; Student perspectives
54	McDonald & Boud (2003)	Experiment	Barbados	High school (Final year)	515	Investigated the effects of self-assessment training on student end-of-high-school qualifications results	Achievement
55	Mac Iver (1987)	Survey	United States	Elementary school (G5 & 6)	1,570	Examined the effects of task structure, ability grouping, emphasis on grades, sex, math talent, and math performance on student self-assessments	Accuracy
56	Miller, Duffy, & Zane (1993)	Experiment	United States	Elementary school (G6)	13	Investigated the effects of rewards on accuracy and achievement on mathematics homework self-correction	Accuracy; Achievement

(Continued)

Table 21.1 (Continued)

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
57	Mitman & Lash (1988)	Survey	United States	Elementary school (G3)	131	Compared student perceptions of academic standing with alternative measures	Accuracy
58	Morrison, Montemayor, & Wiltshire (2004)	Quasi-experiment	United States	Middle & high school	141	Compared student self-assessments of musical performance (with and without models) with expert evaluations of performance	Achievement
59	Munns & Woodward (2006)	Action research	Australia	Elementary school (G4)	Not given	Examined relationships between self-assessment, student engagement, behavior, and student–teacher relationship quality	Self-regulation
60	Ng & Earl (2008)	Survey	Australia	High school (Final year)	94	Investigated the role of feedback, goal orientation, and self-efficacy on student self-estimate accuracy	Accuracy
61	Olina & Sullivan (2002)	Experiment	Latvia	High school	189	Explored the effects of teacher evaluation and student self-evaluation on student performance and efficacy attitudes	Achievement; Self-regulation
62	Peterson & Irving (2008)	Focus Group interviews	New Zealand	High school (G9 & 10)	41	Investigated student conceptions of assessment and feedback	Student perspectives
63	Pomerantz & Ruble (1997)	Survey	United States	Elementary school (G2–5)	236	Examined student conceptions of ability and their relationships to self-evaluation	Accuracy
64	Powel & Gray (1995)	Experiment	United States	Elementary school (G1)	124	Investigated if rewards or collaboration with peers improved the accuracy of student predictions of task success.	Accuracy
65	Raider-Roth (2005)	Interview	United States	Elementary school (G6)	9	Examined students' understandings of self-assessment and their experiences with specific self-assessment tasks	Student perspectives
66	Ramdass & Zimmerman (2008)	Experiment	United States	Elementary school (G5–6)	42	Explored relationships between self-correction training, self-efficacy, self-evaluation, and math performance	Self-regulation; Achievement; Accuracy

	Source	Design	Country	School Level	N	Description of Study	Theme(s)
67	Ross, Hogaboam-Gray, & Rolheiser (2002)	Experiment	Canada	Elementary school (G5–6)	516	Studied the effects of self-evaluation training on mathematics achievement	Achievement
68	Ross, Rolheiser, & Hogaboam-Gray (1998a)	Experiment	Canada	Elementary school (G5–6)	300	Explored the effects of self-assessment training on student mathematics achievement	Achievement; Student perspectives
69	Ross, Rolheiser, & Hogaboam-Gray (1998b)	Quasi-experiment	Canada	Elementary & high school (G5–12)	368	Examined student perspectives of self-assessment when classes were in action research and skills training conditions	Student perspectives
70	Ross, Rolheiser, & Hogaboam-Gray (1999)	Experiment	Canada	Elementary school (G5–6)	296	Evaluated the effects of self-evaluation of narrative writing with rubrics on achievement	Achievement
71	Ross, Rolheiser, & Hogaboam-Gray (2002)	Interviews	Canada	Elementary school (G2, 4, 6)	71	Investigated how student cognition mediated evaluation and achievement	Achievement; Student perspectives; Accuracy; Self-regulation
72	Sadler & Good (2006)	Experiment	United States	Middle school (G7)	126	Examined the accuracy and effects on achievement of students grading their own work using rubrics	Achievement; Accuracy
73	Schunk (1996)	Experiment	United States	Elementary school (G4)	44	Investigated how goals and self-evaluation affect motivation and achievement outcomes	Achievement; Self-regulation
74	Spaights (1965)	Survey	United States	Middle school (G7)	80	Examined self-ratings of ability for reading, arithmetic, and language	Accuracy
75	Stipek (1981)	Interview	United States	Elementary school (K–3)	64	Compared students', teachers', and peers' self-ratings of student academic ability	Accuracy
76	Stipek & Tannatt (1984)	Interview	United States	Elementary (pre, K–3) &	96	Investigated how students judge personal and peers' academic ability	Accuracy

(Continued)

Table 21.1 (Continued)

	Source	*Design*	*Country*	*School Level*	*N*	*Description of Study*	*Theme(s)*
77	Sung, Chang, Chang, & Yu (2010)	Survey	Taiwan	Middle school (G7, 8)	226	Investigated self- and peer assessments rating behaviors and examining reliability and validity	Accuracy
78	van Kraayenoord & Paris (1997)	Survey	Australia	Elementary school (G4–6)	93	Tested a classroom interview instrument designed to assess and promote learning self-appraisal	Accuracy; Achievement
79	Wall (1982)	Experiment	United States	Elementary (G4)	85	Investigated the relative effects of systematic self-monitoring and self-reinforcement on children's academic test performances	Achievement
80	Wan-a-rom (2010)	Case Study	Thailand	High school (G11)	5	Evaluated the accuracy of student English vocabulary knowledge self-assessments	Accuracy
81	Watt (2000)	Survey	Australia	Middle school (G7)	400	Investigated change in student self- and task evaluations, and achievement behavior in mathematics and English	Accuracy
82	Wells & Sweeney (1986)	Longitudinal survey	USA	High school (G10 & 11)	1,508	Examined the relationship between self-esteem and self-assessment of ability	Accuracy
83	Wilson & Wright (1993)	Survey	United States	Middle & high school (G8–12)	301	Investigated if student self-evaluations, grades, and teacher assessments predicted standardized test scores	Accuracy
84	Wright & Houck (1995)	Survey	United States	High school (G9–12)	222	Examined gender differences in self-assessments, teacher ratings, and performance on verbal and numerical reasoning tests	Accuracy

Table 21.1 Summary Design, Demographic, and Thematic Characteristics of Reviewed Studies on Student Self-Evaluation

Relationship of Self-Assessment to Academic Achievement

A number of studies have shown that students who engage in self-assessment experience positive gains in their learning (Table 21.2). While most studies report positive effects of having students self-assess, some reported nil to small effects (i.e., $d \leq 0.20$). The median effect lies between 0.40 and 0.45, a moderate effect consistent with values reported in Black and Wiliam (1998).

Training in diverse self-assessment strategies led to learning gains. For example, immediate self-correction of spelling words generated improved test score performances among primary students (Harward, Allred, & Sudweeks, 1994). Mathematics performance was boosted through the classroom implementation of self-assessment strategies (Ross, Hogaboam-Gray, & Rolheiser, 2002), and students taught self-correction strategies for mathematical long division outperformed the control group (Ramdass & Zimmerman, 2008). Students supported in self-questioning their writing with a computerized prompt system had statistically significant advantages in revision quality (Daiute & Kruidenier, 1985). A 12-month training program in the use of self-assessment processes resulted in a statistically significant advantage to students in high school qualifications examinations (McDonald & Boud, 2003). Being taught explicitly to self-regulate their writing processes resulted in both better writing outcomes and more optimistic self-efficacy and ability self-evaluation (Glaser, Kessler, Palm, & Brunstein, 2010). Perhaps the small effects found in Andrade and Boulay (2003) are attributable to the lack of training in self-assessment students received prior to the study.

Using models, answers, or teacher feedback to guide self-assessment judgments also generally improved performance. Self-rating one's own music performance in conjunction with listening to a model performance improved actual performance (Hewitt, 2001). Self-evaluation combined with teacher evaluation produced better quality science project reports than no-evaluation or teacher-only evaluation, though not better test scores (Olina & Sullivan, 2002).

Children who self-evaluated in conjunction with defining criteria and receiving feedback from teachers about their self-evaluations had small gains in narrative writing (Ross, Rolheiser, & Hogaboam-Gray, 1999). However, large gains were reported in writing for a rubric-guided self-evaluation without teacher feedback (Andrade et al., 2008; Andrade et al., 2010). Similarly, science students who self-graded their work with a rubric that they had co-constructed with their teachers gained considerably more on a teacher-marked science test than students who engaged in peer marking (Sadler & Good, 2006), with much larger gains seen among the initially lower performing students. Ross et al. (1999) also found lower achieving students gained considerably ($d = 0.58$) from being taught to self-assess.

Systems where students predicted or monitored their accuracy and achievement and/or rewarded themselves for accuracy or improvement also were correlated with gains. Self-monitoring the number of answers correct and setting stringent performance standards with self-selected rewards for meeting those standards improved learning of vocabulary and mathematics (Barling, 1980). Students taught to give themselves rewards for reaching challenging targets had modest improvements in achievement when they self-corrected their mathematics homework (Miller, Duffy, & Zane, 1993). Likewise, self-determined reinforcement (i.e., giving themselves rewards based on targets relative to previous performances) gave large learning gains relative to just self-marking (Wall, 1982). Schunk (1996) found that when students were asked to self-assess their ability to accurately complete fraction problems, performance goal orientation resulted in greater effects than learning goal orientation, perhaps because students responded positively to the challenge of getting more problems done and solved. Koivula, Hassmén, & Hunt (2001) found that students who were asked to self-assess the accuracy of their responses to particular standardized test items scored better than pupils who did not take part in this additional monitoring and reflection.

Hence, it appears that there is empirical evidence that self-assessment of a task or self-confidence in the quality of the work will generally improve academic performance across a range of grade levels and subject areas, although the extent of these gains varies across studies, with 11 of the 24 effects falling below the 0.40 Hattie (2009) recommends as a cut score for determining if an intervention is academically

Study	*Type of Self-Assessment*	*Effect size (Cohen's d)*
Wall (1982)	Self-marking with self-selected reinforcements	1.62
Ramdass & Zimmerman (2008)	Self-rated confidence in accuracy of own work	1.50
Schunk (1996)	Self-rated confidence in accuracy of own work (performance goal condition)	1.40
Andrade, Du, & Wang (2008)	Rubric guided judgment	0.87
Sadler & Good (2006)	Rubric guided judgment	0.82
van Kraayenoord & Paris (1997)	Student verbal self-assessments evaluated by researchers	0.77
Andrade, Du, & Mycek (2010)	Rubric guided judgment	0.66
Hewitt (2001)	Self-rated performance	0.59
Olina & Sullivan (2002)	Self-rated written work	0.57
Daiute & Kruidenier (1985)	Computer assisted monitoring of work	0.52
McDonald & Boud (2003)	Monitoring of self-regulation processes	0.45
Ross, Hogaboam-Gray, & Rolheiser (2002)	Generic self-assessment of mathematics	0.40
Glaser et al. (2010)	Self-evaluation of written work	0.38
Schunk (1996)	Self-rated confidence in accuracy of own work (learning goal condition)	0.38
Miller, Duffy, & Zane (1993)	Self-correction of homework	0.32
Koivula, Hassmén, & Hunt (2001)	Self-rated confidence in accuracy of quantitative work	0.29
Barling (1980)	Self-monitoring of accuracy with self-selected rewards and standards	0.28
Harward, Allred, & Sudweeks (1994)	Immediate self-correction of test performance	0.27
Ross, Rolheiser, & Hogaboam-Gray (1999)	Rubric guided judgment	0.18
Koivula, Hassmén, & Hunt (2001)	Self-rated confidence in accuracy of verbal work	0.12
Ross, Rolheiser, & Hogaboam-Gray (1998a)	Self-assessment survey rating of performance and strategy usage on a mathematics test	0.08
Andrade & Boulay (2003)	Rubric guided judgment (response to literature essay)	0.04
Andrade & Boulay (2003)	Rubric guided judgment (historical fiction essay)	–0.04

Table 21.2 Effect Sizes for Learning Effects of Self-Evaluation

worthwhile. These findings also reinforce the claim that it is the implementation and complexity of the self-assessment, more so than the type, which generates the positive effects. While studies using rubrics account for some of the higher effect sizes, three of the lowest effect sizes were also of this type, although the two lowest effects occurred in a study where students used rubrics without any training.

Effect of Self-Assessment on Self-Regulation Processes

Studies have demonstrated that engagement in self-assessment also contributes to increased self-regulating skills (Klenowski, 1995; Ramdass & Zimmerman, 2008), a demonstrated precursor of improved achievement (Schunk, 2005). From self-assessment, greater internality of control (Fernandes & Fontana, 1996) and greater self-focused comparison rather than comparing to peer performance (Ross, Rolheiser, & Hogaboam-Gray, 2002) have been reported (see also Chapter 3 of this volume). Greater persistence on a difficult task was found after confidential self-evaluation of performance in word spelling (Hughes, Sullivan, & Mosley, 1985). Through self-assessment, students thought about their use of strategies for memorizing and recalling mathematics facts, instead of just using rote learning (Brookhart et al., 2004).

Improved student motivation, self-efficacy, engagement, student behavior, and quality of student–teacher relationships have all been found as a consequence of self-evaluation (Glaser et al., 2010; Griffiths & Davies, 1993; Munns & Woodward, 2006; Olina & Sullivan, 2002; Schunk, 1996). Student goal setting, a self-regulating skill connected to self-evaluation, was not a statistically significant factor in improved reading performance and motivation—perhaps because students found it difficult to decide on appropriate, challenging goals (McDevitt, et al., 2008). The effects of self-assessment by student sex have not been extensively studied and are varied. Andrade, Wang, Du, and Akawi (2009) found that while mean student self-reported self-efficacy scores generally increased when using rubrics and self-assessment during the writing process, girls appeared to gain more self-efficacy from the self-assessment training than boys. Frey and Ruble (1987) found girls made more negative self-evaluations and attributions than boys—perhaps because of their concern to maintain social relationships. In contrast, McDonald (2009) found that male students especially benefitted from self-assessment training in relation to motivation and achievement. However, Johnson and Winterbottom (2011) found that students in the girls-only class they studied reported lower motivation, lower commitment to a mastery goal orientation, and lower self-efficacy after the implementation of self- and peer assessment, although observed class behaviors suggested some students became more learning oriented.

The research evidence for the connection between self-assessment and self-regulated learning (SRL) is not robust, despite many assertions to that effect. While evidence tentatively appears to suggest that self-assessment can positively contribute to student motivation and self-regulation, some results are mixed. It remains unclear which particular types of students may benefit the most from these practices as it is likely that pupils have highly individualized responses to self-assessment, as discussed in the next section.

Student Perceptions of Self-Assessment

Some studies indicate students seem to enjoy being involved in self-assessment (Brookhart et al., 2004; McDonald, 2009; Ross, Rolheiser, et al., 2002), especially if self-assessment helps them improve their understanding of criteria or work toward their own goals (McDevitt, et al., 2008). Reviews note that rubrics have been found to be particularly helpful for getting students to better understand evaluative criteria (Andrade, 2000; Andrade & Valtcheva, 2009).

Notwithstanding these demonstrated effects, a number of studies have shown that many students raise questions about self-assessment. Students are not always positive about self-assessment or aware of what it is really for. McDonald (2002) found that the students defined self-assessment primarily in terms of autonomous study skills rather than reflections on or evaluations of the merit of their own work, although students in her later study described self-assessment as helpful and motivating (McDonald, 2009). Sometimes students simply fill in the blanks rather than engage in thoughtful self-evaluation (Brookhart et al., 2004). There is evidence that students do not always consider self-assessment to even be assessment (Brown, Irving,

Peterson, & Hirschfeld, 2009; Brown, Peterson, & Irving, 2009; Harris, Harnett, & Brown, 2009; Peterson & Irving, 2008) and question its value (LaVoie & Hodapp, 1987), still wanting thorough, individualized teacher feedback (Lasonen, 1995). Students sometimes see self-assessment as boring, an inappropriate appropriation of the teacher's responsibility, and/or a source of cheating or nonstandard scores (Gao, 2009; Harris & Brown, 2010; Johnson & Winterbottom, 2011; Peterson & Irving, 2008; Ross et al., 1998b). Ross et al. (1998b)found that teachers did little to explore student misconceptions and concerns about self-assessment, leading many pupils to become increasingly negative.

Students have also raised concerns about their psychological safety when their self-evaluations are made public to peers, parents, and teachers (Cowie, 2009; Harris & Brown, 2010; Raider-Roth, 2005; Ross, Rolheiser, et al., 1998b, 2002), a common classroom process (Kasanen & Räty, 2002). Consequently, students may provide depressed self-evaluations for fear of being seen as egotistical (Brooks, 2002) or for cultural practices such as self-effacement (Kwok & Lai, 1993). Alternatively, they may give elevated self-assessments to avoid being shamed in front of the class (Harris & Brown, 2010), with studies showing students have differing and highly personal reactions to self-assessment disclosure (Cowie, 2009; Harris et al., 2009). Hence, if self-assessment is to be an effective classroom practice, the valid concerns students have about its legitimacy and practice must be taken into account.

Accuracy in Self-Assessment

Studies reviewed by Ross (2006) indicate that the student as a self can be highly consistent in evaluations, but comparisons between self-evaluations and other measures (e.g., test scores, teacher ratings, and parent ratings) depict a less reliable portrait for self-assessment. The correlation between self-ratings and teacher ratings (Alsaker, 1989; Connell & Ilardi, 1987; Sung, Chang, Chang, & Yu, 2010; van Kraayenoord & Paris, 1997), between self-estimates of performance and actual test scores (Ash, 1980; Barnett & Hixon, 1997; Bradshaw, 2001; Ikeguchi, 1996; Koivula et al., 2001; LaVoie & Hodapp, 1987; Luyten & Dolkar, 2010; Wilson & Wright, 1993; Wright & Houck, 1995), and between student and teacher rubric-based judgments (Higgins, Harris, & Kuehn, 1994; Laveault & Miles, 2002; Sadler & Good, 2006) tended to be positive, ranging from weak to moderate (i.e., values ranging from $r \approx 0.20$ to 0.80), with few studies reporting correlations greater than 0.60. Accuracy was improved when students were taught explicitly to use a self-checking strategy (Ramdass & Zimmerman, 2008), and rewarding accuracy was also found to increase it (Miller et al., 1993). Nonetheless, the accuracy of student self-assessment does not appear to be uniform throughout the student's life course, nor across the full range of learning activities. Some students do not accept that their assessments are inherently less accurate than teachers, believing self-assessments should be used for grading purposes (Chang & Tseng, 2011).

Accuracy, Age, and Schooling Experience

Increasing age is confounded with increasing experience of school so it is not entirely clear whether improved accuracy of self-evaluation is a function of developmental processes or educational experience. Nonetheless, younger children tend to be more optimistic in their self-estimations of performance than older children (Frey & Ruble, 1987; Eccles, Wigfield, Harold, & Blumenfeld, 1993; Ross, Rolheiser, et al., 2002). A review by Stipek and Mac Iver (1989) noted that in elementary school, the criteria children use to judge their intellectual competence starts with emphasis on effort, social reinforcement, and mastery, maturing to a reliance on more objective and normative information.

In studies that used self-ratings, younger students tend to be more optimistic, lenient, or generous than older students (Blatchford, 1997a, 1997b; Kaderavek, Gillam, Ukrainetz, Justice, & Eisenberg, 2004; Kasanen, Räty, & Eklund, 2009; Stipek, 1981; Stipek & Tannatt, 1984; Wilson & Wright, 1993). Elder (2010) found that Grade 1 students reported focusing on superficial features, while Grades 4 and 5 students described making more complex judgments; however, both groups indicated relying heavily on the opinions of others (e.g., parents or teachers) when making decisions about work quality. Older students' self-ratings, while lower than younger students, tend to correlate more strongly with teacher ratings or test scores (Alsaker, 1989; Blatchford, 1997a; Bradshaw,

2001; Butler, 1990; Hewitt, 2005; Kaderavek et al., 2004; Pomerantz & Ruble, 1997; Stipek, 1981; Stipek & Tannatt, 1984) and are generally more sophisticated (Ross, Rolheiser, et al., 2002).

Accuracy and Academic Ability

A large number of studies suggest that accuracy in self-assessment is related to academic ability—that is, higher performing students evaluate their own work more accurately. Only one study was found that contradicted this pattern (Spaights, 1965); although, given the small sample size and its age, the findings may not give an accurate picture of the current situation. Consistent with the notion of double-handicapping related to low ability, high ability students seem to be more severe in assessing their work than their teachers, while low ability students seem to be more lenient on themselves (Barnett & Hixon, 1997; Claes & Salame, 1975; Kwok & Lai, 1993; Laveault & Miles, 2002; Mitman & Lash, 1988; Sung et al., 2010; Watt, 2000). The self-ratings from more able, proficient, or intelligent students tend to correlate more highly with teacher and test measures than the ratings of less proficient students (Claes & Salame, 1975; Keil, McClintock, Kramer, & Platow, 1990; Kwok & Lai, 1993; Laveault & Miles, 2002; Mitman & Lash, 1988; Ng & Earl, 2008; Sung et al., 2010; van Kraayenoord & Paris, 1997). This may suggest that the path to improved performance is not through inflated but inaccurate confidence in one's ability, but through greater humility due to one's appreciation of competence and capability. Hence, empirical data show that age and proficiency are a powerful basis for more accurate self-evaluation.

Accuracy and Task Difficulty

The difficulty of the task being learned interacts with students' ability to self-assess (Barnett & Hixon, 1997; Bradshaw, 2001; Hewitt, 2005). Tasks that are familiar and predictable probably permit more accurate student self-assessment. More technically difficult tasks require greater attention and effort, and this probably interferes with resources needed to monitor and self-rate performance. For example, the simple, concrete task of evaluating the accuracy of letter formation had high levels of agreement between student self-scoring and teacher scoring (Jones, Trap, & Cooper, 1977), and students were 80% to 90% accurate in their self-assessments of whether or not they knew the meaning of a word (Wan-a-rom, 2010). However, Powel and Gray (1995) could not reduce the inaccuracy of young students' self-estimates of success in a beanbag tossing task, despite its obvious concrete nature. Self-assessment in hard tasks can be supported with extra performance-based feedback (Lee & Gavine, 2003). The presence or absence of formal instruction in tested content prior to testing appears to impact student ability to predict accurately their performance (Barnett & Hixon, 1997), and greater accuracy in self-assessment was found when it was explicitly linked to an assessment of the same proficiency (Butler & Lee, 2006).

Basis for Evaluation

Studies have indicated that students value and use criteria based on construct irrelevant factors like effort when evaluating their work (e.g., Ross, Rolheiser, et al., 1998b, 2002). However, self-assessments that use more specific, concrete standards or reference points, rather than subjective criteria (e.g., "I made an effort" or "I'm good at this"), are associated with greater accuracy (Claes & Salame, 1975). Students who received regular teacher feedback in math were found to be more accurate in their self-assessments as they were more likely to use legitimate criteria to judge their abilities (Mac Iver, 1987). More modest and more accurate self-assessments were found among older students who shifted from a general social comparison (i.e., "all children my age") to a more specific social comparison (i.e., "those in my class") as the basis for self-rating (Blatchford, 1997a).

Other Factors Related to Accuracy

It has also been reported that gender, ethnic culture, and personality impact accuracy. For example, Blatchford (1997b) found that as students grew older, White students (especially girls) were less positive and less accurate in their self-assessments of academic achievement than Black students. Wells and Sweeney (1986) identified that students with consistently high self-esteem were more likely to overestimate their abilities, while those with low self-esteem often underestimated their abilities.

Training is also likely to improve accuracy. For example, improved accuracy in rubric-based self-assessment has been demonstrated (1) by teaching students to use explicit, objective criteria (Ramdass & Zimmerman, 2008); (2) by involving students in the co-construction of criteria for the rubric and with practice at using the rubric (Ross, Rolheiser, & Hogaboam-Gray, 1998a); (3) by ensuring students are motivated to pay attention to the rubric (Laveault & Miles, 2002); and (4) by getting students to justify their self-evaluation explicitly to their peers (Dunning et al., 2004).

Summary of Accuracy in Self-Assessment

The general impression formed from the research is that self-assessment is not robustly accurate but also it certainly is not randomly related to external measures of performance. Correlations falling in the range of 0.30 to 0.50 explain some 10% to 25% of variance between the self-assessment and some external measure of performance. Student self-assessments appear to be more accurate among older or more academically able students. Furthermore, students tend to assign lower and less optimistic ratings to their own work with increased experience or ability. Underrating of ability, found in older and more able students, was also correlated with less anxiety and less "emotional investment in achievement outcomes" (Connell & Ilardi, 1987, p. 1303). Hence, as students mature and develop academically, we can expect self-assessments to become less optimistic and more accurate. Educators should not panic when students begin to assign lower ratings for their own work as this may indicate improved competence and a more accurate self-evaluation of performance. While training in self-assessment can improve the accuracy of self-assessment, it seems pedagogically inappropriate to encourage high self-assessment scores independent of increased academic competence; students should not be encouraged to go easy on themselves for ego protection purposes.

Nonetheless, there is a need for instructional input and caution when implementing self-assessment with students likely to be relatively inaccurate (i.e., younger or less proficient students). All self-assessment techniques seem to have similar ranges of agreement with external measures, and rubric-based self-assessment studies appear most promising because of the relatively high learning effects shown when students use them. The studies reviewed also point to the importance of reducing the subjectivity in the criteria students use to evaluate their work. The provision of rubrics and a focus on what others would deem as quality appear to be necessary for high quality self-assessment. Concern must be expressed about the wisdom of using student self-assessments as part of course grades or final summary evaluations because this introduces high-stakes consequences for honest, accurate evaluations.

Conclusion

The reviewed studies suggest that student self-assessment can contribute to improved learning outcomes and better self-regulation skills, provided such self-evaluation involves deep engagement with the processes affiliated with self-regulation (i.e., goal setting, self-monitoring, and evaluation against valid, objective standards). It would appear that it is not the form of self-assessment that matters per se but rather the level of mental engagement students must use to determine how well they have done. Low levels of cognitive engagement can be seen in self-rating satisfaction with a happy or smiley face scale, awarding oneself a grade for a test based on perceived effort, or assigning a rubric characteristic based on a desire to avoid failure. Higher levels of self-assessment cognitive engagement can be seen when students rate themselves relative to challenging goals, evaluate test performance on objective criteria, or use rubrics to which they contributed. Learning and self-regulation gains seem to depend on higher levels of mental involvement in the process of determining the quality of work.

However, as predicted by psychometric and psychological theorization, data suggest that school children are usually not very good at this type of critical, metacognitive reflection unless the accuracy factors identified in this chapter are present, making the use of student self-assessment for grading purposes ill-advised. Improved accuracy appears to be partly a function of cognitive developmental processes (i.e., increasing age) and educational practices (i.e., increasing school experience). Additionally, it appears possible to train students to engage in these deep reflective

practices and that such training is associated with better self-regulation of learning, more accurate self-evaluation, and better learning outcomes (Daiute & Kruidenier, 1985; Glaser et al., 2010; McDonald & Boud, 2003; Miller et al., 1993; Morrison, Montemayor, & Wiltshire, 2004; Ramdass & Zimmerman, 2008; Ross et al., 1998a).

Additionally, the teacher clearly has to play an active part in the development and monitoring of self-evaluation, most especially for students who have low academic performance. That low performing students, given they are generally weaker at accurate self-assessment, seem to gain more from this type of self-evaluative reflection is especially good news for educators, as closing the distance between the best and lowest performers is an important goal of schooling. It would appear that, while better students can already self-evaluate effectively, lower performing students need input (i.e., instruction and feedback) to master this key self-regulatory process. Nonetheless, the involvement of teachers in student self-evaluation shifts the ground from a purely personal experience to a shared, public space in which psychological safety, and trust must be present for students to be capable of producing genuine, honest, and accurate self-assessment. Hence, a cautious seal of approval can be given to the use of the best forms of self-assessment in environments that support good teacher–student rapport.

Implications for Current Pedagogical Practices

This review makes it clear that high quality student self-assessment requires active involvement of both students and teachers; self-assessment/evaluation is not an excuse for teacher absence. Ross (2006) provided four essential techniques that need to be incorporated into CA practices: (1) Students need to be involved in the process of establishing criteria for evaluating work outcomes; (2) students need to be taught how to apply those criteria; (3) feedback from others (i.e., teachers and peers) is needed so that students can move from inaccurate, false self-perceptions of their work to more accurate comprehension of the quality of their work; and (4) students need to be taught how to use other assessment data (e.g., test scores or graded work) to improve their work. To extend this list, we would add a fifth condition: There must be psychological safety in the implementation of self-evaluation. Children must know that it is safe to disclose low performance and that they do not need to resort to score-enhancement strategies.

Implications for Future Research

As Barnett and Hixon (1997) pointed out, it is unclear if the association between accurate self-assessment and higher achievement, while consistent with self-regulation models of learning, is a consequence of improved self-regulation or is a by-product of higher achievement. Thus, more studies are needed to determine conditions under which self-evaluation accuracy can be successfully taught to lower performing students, consequently bringing about higher academic performance. Likewise, self-regulation of learning studies (Ramdass & Zimmerman, 2008; Zimmerman, Bonner, & Kovach, 1996) suggests that low performing students can learn to self-regulate, but the generalizability of those studies, given the constraints on accuracy of self-evaluation identified in this chapter, is still in doubt. Hence, studies are needed to explicitly explore the relationships among self-regulation, self-assessment, and academic achievement.

Additionally, questions are raised about the abilities of young students to accurately self-assess. More research is needed to establish if there is a chronological or developmental age beneath which there is little benefit to be reaped through self-assessment. There is clear evidence that low performing students are most inaccurate in their self-assessments, but several studies have shown that the greatest improvement in performance through self-assessment was seen among the low performing students. This suggests that with training in self-assessment accuracy, the gap between low and high performing students might close. Furthermore, it is unknown if there is an interaction between age and academic ability as factors influencing the accuracy of self-assessment judgments.

Psychological safety within classrooms and across cultures is another factor to consider. More research is required to determine if there are some cultures that are more or less able to adopt accurate self-assessment practices due to socially held beliefs about the self, performance, and

others. It is also worth investigating classroom environmental factors that make students more or less likely to create and disclose accurate self-assessments. The current studies concerning psychological safety all depend on small-scale narrative techniques; studies are needed that establish in a more generalizable way how student psychological safety can be achieved and whether it improves the quality of self-assessment. How student personality factors mediate self-assessment is also unknown.

A developmental process (e.g., Piagetian development of abstract cognitive reasoning) or an experiential process seems to underlie the general phenomenon of increased accuracy with age. Stipek, Recchia, and McClintic (1992) proposed an empirically derived developmental sequence for reactions to achievement situations in preschool children (ages 1–5) in which children seek positive reactions and avoid potential negative reactions from adults prior to developing a more independent evaluation. However, the inaccuracy of school children's self-estimates suggests that considerable maturation is needed before improvements can be detected (Powel & Gray, 1995). While there may be a developmental trend in accuracy of self-assessment, Alsaker (1989) correctly identified that longitudinal studies are needed before firm conclusions about the underlying processes can be drawn. It is possible that increasing knowledge rather than cognitive or emotional development is sufficient to improve the quality of self-evaluations; the research to date appears insufficient to answer this question. Burnett (1996) rightly pointed out that the decline in self-concept evaluation associated with increasing age may be a function of schooling extinguishing student optimism rather than the development of greater realism; it remains to be seen how this could be tested given ethical and practical constraints about manipulating schooling processes.

Not addressed in this review is a deep analysis of the various techniques of self-assessment. How and when self-evaluations of proficiency, competence, or performance are obtained is still highly variable. Studies have collected self-evaluations before and immediately after assessment events and prior to instruction. Self-estimation of performance has used norm-referencing, absolute referencing, grades and scores, self-centered evaluations, and estimations in terms of the objective criteria of a rubric. The general trend seems to be that the more concrete and immediate the evaluation is and the more the student is cognitively engaged in evaluating quality characteristics, the greater the likelihood that students will make an accurate assessment of their capabilities and improve their learning. Positive effects on learning and self-regulation were seen through self-evaluation techniques that moved most strongly away from simple self-marking or self-rating. Consistent with arguments about the need for metacognitive involvement in self-evaluation (Zimmerman, 2002), these studies show that depth of processing and engagement in self-assessment is required for it to have a learning effect. However, there is no definitive gold standard method for helping students evaluate their own work.

Research into improving the quality of our methods of data collection for research purposes, let alone educational application, is still warranted. It would be useful for large-scale experimental studies to examine which modes of self-assessment allow students to create the most accurate judgments and which, if any, lead to improved motivation, psychological safety, self-regulation, and academic performance over the longer term. Studies that identify the type of learning arising from each method of self-assessment also appear warranted. The trend seems to be that self-assessments that require high levels of cognitive involvement have the greatest learning effects, though it is possible that this is a practice effect from frequent self-assessment rather than a self-regulatory process.

Furthermore, research into the consequential validity of self-evaluations is warranted. When students self-assess and get it wrong or right, what do they do with that information? What are the low- and high-stakes consequences of accurate and inaccurate student self-assessments? While accuracy would appear to be essential, it may be that inaccurate self-assessment in classroom settings—where teachers can structure learning environments and activities—has little negative impact on a student. While this seems improbable to us, the current research literature does not appear to definitively address this problem. Hence, future studies that examine in detail what students do with their self-evaluations—especially when they are palpably wrong—is of great importance to improving our understanding of student self-evaluation.

References

Alsaker, F. D. (1989). School achievement, perceived academic competence and global self-esteem. *School Psychology International, 10*(2), 147–158.

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 1–7.

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

Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). New York: Routledge.

Andrade, H. L., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice, 17*(2), 199–214.

Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice, 17*(2), 3–13.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice, 28*(1), 12–19.

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *Journal of Educational Research, 102*(4), 287–302.

Ash, R. A. (1980). Self-assessments of five types of typing ability. *Personnel Psychology, 33*(2), 273–281.

Barling, J. (1980). A multistage, multidependent variable assessment of children's self-regulation of academic performance. *Child Behavior Therapy, 2*(2), 43–54.

Barnett, J. E., & Hixon, J. E. (1997). Effects of grade level and subject on student test score predictions. *Journal of Educational Research, 90*(3), 170–174.

Black, P., & Harrison, C. (2001). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review, 83*(302), 43–49.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Blatchford, P. (1997a). 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), 345–359.

Blatchford, P. (1997b). Pupils' self assessments of academic attainment at 7, 11 and 16 years: Effects of sex and ethnic group. *British Journal of Educational Psychology, 67*(2), 169–184.

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

Bradshaw, B. K. (2001). Do students effectively monitor their comprehension? *Reading Horizons, 41*(3), 143–154.

Brookhart, S. M. (2009). *Grading*. Upper Saddle River, NJ: Pearson.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57,* 213–227.

Brooks, V. (2002). *Assessment in secondary schools: The new teacher's guide to monitoring, assessment, recording, reporting and accountability*. Buckingham, UK: Open University Press.

Brown, G. T. L. (2009). The reliability of essay scores: The necessity of rubrics and moderation. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment and higher education student outcomes: Policy, practice and research* (pp. 40–48). Wellington, New Zealand: Ako Aotearoa.

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning & Instruction, 19*(2), 97–111.

Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Self-regulatory beliefs about assessment predict mathematics achievement. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 159–186). Charlotte, NC: Information Age Publishing.

Burnett, P. C. (1996). Gender and grade differences in elementary school childrens' descriptive and evaluative self-statements and self-esteem. *School Psychology International, 17*(2), 159–170.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281.

Butler, R. (1990). The effects of mastery and competitive conditions on self-assessment at different ages. *Child Development, 61*(1), 201–210.

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *Modern Language Journal, 90*(4), 506–518.

Byrne, B. M., & Bazana, P. (1996). Investigating the measurement of social and academic competencies for early/late preadolescents and adolescents: A multitrait-multimethod analysis. *Applied Measurement in Education, 9*(2), 113–132.

Chang, C. C., & Tseng, K. H. (2011). Using a web-based portfolio assessment system to elevate project-based learning performances. *Interactive Learning Environments, 19*(3), 211–230.

Claes, M., & Salame, R. (1975). Motivation toward accomplishment and the self-evaluation of performances in relation to school achievement. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 7*(4), 397–410.

Clarke, S. (2005). *Formative assessment in the secondary classroom.* Abingdon, UK: Hodder Murray.

Clarke, S., Timperley, H. S., & Hattie, J. A. (2003). *Unlocking formative assessment: Practical strategies for enhancing students' learning in the primary and intermediate classroom* (New Zealand ed.). Auckland, New Zealand: Hodder Moa Beckett Publishers Limited.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Connell, J. P., & Ilardi, B. C. (1987). Self-system concomitants of discrepancies between children's and teachers' evaluations of academic competence. *Child Development, 58*(5), 1297–1307.

Cowie, B. (2009). My teacher and my friends helped me learn: Student perceptions and experiences of classroom assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 85–105). Charlotte, NC: Information Age Publishing.

Daiute, C., & Kruidenier, J. (1985). A self-questioning strategy to increase young writers' revising processes. *Applied Psycholinguistics, 6*(3), 307–318.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer-, and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*(3), 69–106.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development, 64*(3), 830–847.

Educational Research Information Center (U.S.). (2001). *Thesaurus of ERIC descriptors* (14th ed.). Westport, CT: The Oryx Press. Retrieved from www.eric.ed.gov/ERICWebPortal/resources/html/thesaurus/about_thesaurus.html

Elder, A. D. (2010). Children's self-assessment of their school work in elementary school. *Education, 38*(1), 5–11.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430.

Fernandes, M., & Fontana, D. (1996). Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. *British Journal of Educational Psychology, 66*(3), 301–313.

Frey, K. S., & Ruble, D. N. (1987). What children say about classroom performance: Sex and grade differences in perceived competence. *Child Development, 58*(4), 1066–1078.

Gao, M. (2009). Students' voices in school-based assessment of Hong Kong: A case study. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 107–130). Charlotte, NC: Information Age Publishing.

Glaser, C., Kessler, C., Palm, D., & Brunstein, J. C. (2010). Improving fourth graders' self-regulated writing skills: Specialized and shared effects of process-oriented and outcome- related self-regulation procedures on students' task performance, strategy use, and self-evaluation. *Zeitschrift fur PadagogischePsychologie/ German Journal of Educational Psychology, 24*(3-4), 177–190.

Griffiths, M., & Davies, C. (1993). Learning to learn: Action research from an equal opportunities perspective in a junior school. *British Educational Research Journal, 19*(1), 43–58.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Westport, CT: Praeger.

Harris, L. R., & Brown, G. T. L. (2010, May). "*My teacher's judgment matters more than mine*"*: Comparing teacher and student perspectives on self-assessment practices in the classroom.* Paper presented to the SIG-Classroom Assessment at the American Educational Research Association Annual Conference, Denver, CO.

Harris, L. R., Harnett, J. A., & Brown, G. T. L. (2009). 'Drawing' out student conceptions: Using pupils' pictures to examine their conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 321–330). Charlotte, NC: Information Age Publishing.

Harward, S. V., Allred, R. A., & Sudweeks, R. R. (1994). The effectiveness of our self-corrected spelling test methods. *Reading Psychology, 15*(4), 245–271.

Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses in education.* New York: Routledge.

Hattie, J. A., & Brown, G. T. L. (2010). Assessment and evaluation. In C. Rubie-Davies (Ed.), *Educational psychology: Concepts, research and challenges* (pp. 102–117). New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hewitt, M. P. (2001). The effects of modeling, self-evaluation, and self-listening on junior high instrumentalists' music performance and practice attitude. *Journal of Research in Music Education, 49*(4), 307–322.

Hewitt, M. P. (2005). Self-evaluation accuracy among high school and middle school instrumentalists. *Journal of Research in Music Education, 53*(2), 148.

Higgins, K. M., Harris, N. A., & Kuehn, L. L. (1994). Placing assessment into the hands of young children: A study of student-generated criteria and self-assessment. *Educational Assessment, 2*(4), 309–324.

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

Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing, 11*(1), 42–65.

Ikeguchi, C. B. (1996). *Self assessment and ESL competence of Japanese returnees.* Retrieved from http://eric.ed.gov/PDFS/ED399798.pdf

Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students.* Thousand Oaks, CA: Corwin Press.

Johnson, N., & Winterbottom, M. (2011). Supporting girls' motivation in science: A study of peer- and self-assessment in a girls-only class. *Educational Studies, 37*(4), 391–403.

Jones, J. C., Trap, J., & Cooper, J. O. (1977). Technical report: Students' self-recording of manuscript letter strokes. *Journal of Applied Behavior Analysis, 10*(3), 509–514.

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.

Kasanen, K., & Räty, H. (2002). "You be sure now to be honest in your assessment": Teaching and learning self-assessment. *Social Psychology of Education, 5*(4), 313–328.

Kasanen, K., Räty, H., & Eklund, A.-L. (2009). Elementary school pupils' evaluations of the malleability of their academic abilities. *Educational Research, 51*(1), 27–38.

Keil, L. J., McClintock, C. G., Kramer, R., & Platow, M. J. (1990). Children's use of social comparison standards in judging performance and their effects on self-evaluation. *Contemporary Educational Psychology, 15*(1), 75–91.

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

Koivula, N., Hassmén, P., & Hunt, D. P. (2001). Performance on the Swedish Scholastic Aptitude Test: Effects of self-assessment and gender. *Sex Roles: A Journal of Research, 44*(11), 629–645.

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 and review of the literature. *Review of Educational Research, 75*(1), 63–82.

Kwok, D. C., & Lai, D. W. (1993, May). *The self-perception of competence by Canadian and Chinese children.* Paper presented at the annual convention of the Canadian Psychological Association, Montreal, QC.

Lasonen, J. (1995). A case study of student self-assessment in upper secondary education. In J. Lasonen & M.-L. Stenstrom (Eds.), *Contemporary issues of occupational education in Finland* (pp. 199–215). Jyvaskyla, Finland: University of Jyvaskyla, Institute for Educational Research.

Laveault, D., & Miles, C. (2002, April). *The study of individual differences in the utility and validity of rubrics in the learning of writing ability.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

LaVoie, J. C., & Hodapp, A. F. (1987). Children's subjective ratings of their performance on a standardized achievement test. *Journal of School Psychology, 25*(1), 73–80.

Lee, D., & Gavine, D. (2003). Goal-setting and self-assessment in Year 7 students. *Educational Research, 45*(1), 49–59.

Lipnevich, A., & Smith, J. (2008). *Response to assessment feedback: The effects of grades, praise, and source of information* (Research report RR-08-30). Princeton, NJ: Educational Testing Service.

Luyten, H., & Dolkar, D. (2010). School-based assessments in high-stakes examinations in Bhutan: A question of trust? Exploring inconsistencies between external exam scores, school-based assessments, detailed teacher ratings, and student self-ratings. *Educational Research and Evaluation, 16*(5), 421–435.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and

meta-analysis. *Journal of Applied Psychology, 67*(3), 280–296.

Mac Iver, D. (1987). Classroom factors and student characteristics predicting students' use of achievement standards during ability self-assessment. *Child Development, 58*(5), 1258–1271.

Marsh, H. W. (1988). *Self Description Questionnaire: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph*. San Antonio, TX: The Psychological Corporation.

McDevitt, T. M., Sheehan, E. P., Sinco, S. R., Cochran, L. S., Lauer, D., & Starr, N. L. (2008). These are my goals: Academic self-regulation in reading by middle-school students. *Reading Improvement, 45*(3), 115–138.

McDonald, B. (2002). Self assessment skills used by high school students without formal training. *School Psychology International, 23*(4), 416–424.

McDonald, B. (2009). Exploring academic achievement in males trained in self-assessment skills. *Education, 37*(2), 145–157.

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: Principles, Policy & Practice, 10*(2), 209–220.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Miller, T. L., Duffy, S. E., & Zane, T. (1993). Improving the accuracy of self-corrected mathematics homework. *Journal of Educational Research, 86*(3), 184–189.

Mitman, A. L., & Lash, A. A. (1988). Student's perceptions of their academic standing and classroom behavior. *Elementary School Journal, 89*(1), 55–68.

Morrison, S. J., Montemayor, M., & Wiltshire, E. S. (2004). The effect of a recorded model on band students' performance self-evaluations, achievement, and attitude. *Journal of Research in Music Education, 52*(2), 116–129.

Munns, G., & Woodward, H. (2006). Student engagement and student self-assessment: The REAL framework. *Assessment in Education: Principles, Policy and Practice, 13*(2), 193–213.

Ng, J. R., & Earl, J. K. (2008). Accuracy in self-assessment: The role of ability, feedback, self-efficacy and goal orientation. *Australian Journal of Career Development, 17*(3), 39–50.

Olina, Z., & Sullivan, H. J. (2002, April). *Effects of teacher and self-assessment on student performance*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Paulhus, D. (1991). Measurement and control of response bias. In J. Robinson, P. Shaver, & L. Wrightsman (Eds.), *Measures of personality and social psychological attitudes, Measures of social psychological attitudes* (Vol. 1, pp. 17–59). San Diego, CA: Academic Press.

Pausch, R., & Zaslow, J. (2008). *The last lecture*. London: Hodder & Stoughton.

Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction, 18*(3), 238–250.

Pomerantz, E. M., & Ruble, D. N. (1997). Distinguishing multiple dimensions of conceptions of ability: Implications for self-evaluation. *Child Development, 68*(6), 1165–1180.

Popham, W. J. (2000). *Modern educational measurement: Practical guidelines for educational leaders* (6th ed.). Boston: Allyn & Bacon.

Powel, W. D., & Gray, R. (1995). Improving performance predictions by collaboration with peers and rewarding accuracy. *Child Study Journal, 25*(2), 141–154.

Raider-Roth, M. B. (2005). Trusting what you know: Negotiating the relational context of classroom life. *Teachers College Record, 107*(4), 587–628.

Ramdass, D., & Zimmerman, B. J. (2008). Effects of self-correction strategy training on middle school students' self-efficacy, self-evaluation, and mathematics division learning. *Journal of Advanced Academics, 20*(1), 18–41.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation, 11*(10). Retrieved from http://pareonline.net/getvn.asp?v=11&n=10

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in Grade 5–6 mathematics: Effects on problem-solving achievement. *Educational Assessment, 8*(1), 43–59.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998a, April). *Impact of self-evaluation training on mathematics achievement in a cooperative learning environment*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998b). 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., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing, 6*(1), 107–132.

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

Ruff, J. C., & Barris, J. (2009). The sound of one house clapping: The unmannerly doctor as zen

rhetorician. In H. Jacoby (Ed.), *House and philosophy: Everybody lies* (pp. 84–97). New York: John Wiley.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78*(3), 450–462.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1–31.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119–144.

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

Schunk, D. H. (2005). Commentary on self-regulation in school contexts. *Learning & Instruction, 15,* 173–177.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, pp. 39–83). Chicago: Rand McNally.

Spaights, E. (1965). Accuracy of self-estimation of junior high school students. *Journal of Educational Research, 58*(9), 416–419.

Stipek, D. J. (1981). Children's perceptions of their own and their classmates' ability. *Journal of Educational Psychology, 73*(3), 404–410.

Stipek, D., & Mac Iver, D. (1989). Developmental changes in children's assessment of intellectual competence. *Child Development, 60*, 521–538.

Stipek, D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development, 57*(1, Serial No. 226).

Stipek, D. J., & Tannatt, L. M. (1984). Children's judgments of their own and their peers' academic competence. *Journal of Educational Psychology, 76*(1), 75–84.

Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence, 33*(1), 135–145.

Todd, R. W. (2002). Using self-assessment for evaluation. *English Teaching Forum, 40*(1), 16–19.

Topping, K. (2003). Self and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55–87). Dordrecht, NL: Kluwer Academic Publishers.

Towler, L., & Broadfoot, P. (1992). Self-assessment in the primary school. *Educational Review, 44*(2), 137–151.

Tuleya, L. G. (2007). *Thesaurus of psychological index terms.* Washington, DC: American Psychological Association.

van Kraayenoord, C. E., & Paris, S. G. (1997). Australian students' self-appraisal of their work samples and academic progress. *The Elementary School Journal, 97*(5), 523–537.

Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education, 34*(2), 239–255.

Wall, S. M. (1982). Effects of systematic self-monitoring and self-reinforcement in children's management of test performances. *Journal of Psychology, 111*(1), 129–136.

Wan-a-rom, U. (2010). Self-assessment of word knowledge with graded readers: A preliminary study. *Reading in a Foreign Language, 22*(2), 323–338.

Watt, H. M. (2000). Measuring attitudinal change in mathematics and English over the 1st year of junior high school: A multidimensional analysis. *Journal of Experimental Education, 68*(4), 331–361.

Wells, L., & Sweeney, P. D. (1986). A test of three models of bias in self-assessment. *Social Psychology Quarterly, 49*(1), 1–10.

Williams, J. E. (1996, April). *Academic self-concept to performance congruence among able adolescents.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Wilson, D. B. (2001). *Effect size determination program* (Version 2.0) (Excel Macro Application). College Park: University of Maryland.

Wilson, J., & Wright, C. R. (1993). The predictive validity of student self-evaluations, teachers' assessments, and grades for performance on the Verbal Reasoning and Numerical Ability scales of the Differential Aptitude Test for a sample of secondary school students attending rural Appalachia schools. *Educational and Psychological Measurement, 53*(1), 259–270.

Wright, C. R., & Houck, J. W. (1995). Gender differences among self-assessments, teacher ratings, grades, and aptitude test scores for a sample of students attending rural secondary schools. *Educational and Psychological Measurement, 55*(5), 743–752.

Zimmerman, B. J. (1998). Academic studying and the development of personal skill: A self-regulatory perspective. *Educational Psychologist, 33,* 73–86.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy.* Washington, DC: American Psychological Association.