Educational Theory and Practice Faculty Scholarship

Educational Theory and Practice

2009

Rubric-Referenced Self-Assessment and Self-Efficacy for Writing

Heidi Andrade
University at Albany, State University of New York, handrade@albany.edu

Xiaolei Wang
University at Albany, State University of New York

Ying Du
University at Albany, State University of New York

Robin L. Akawi
University at Albany, State University of New York

Rubric-Referenced Self-Assessment and Self-Efficacy for Writing

HEIDI L. ANDRADE
XIAOLEI WANG
YING DU
ROBIN L. AKAWI
University at Albany–State University of New York

ABSTRACT. The authors investigated the relation between long- and short-term rubric use (including self-assessment), gender, and self-efficacy for writing by elementary and middle school students (N = 268). They measured long-term rubric use with a questionnaire. They manipulated short-term rubric use by a treatment that involved reviewing a model and using a rubric to self-assess drafts. The authors collected self-efficacy ratings 3 times. Results revealed that girls' self-efficacy was higher than boys' self-efficacy before they began writing. The authors found interactions between gender and rubric use: Average self-efficacy ratings increased as students wrote, regardless of condition, but the increase in the self-efficacy of girls in the treatment group was larger than that for girls in the comparison group, and long-term rubric use associated only with the self-efficacy of girls.

Keywords: gender, rubric, self-assessment, self-efficacy, writing

Rubrics have become popular with teachers as a means of communicating expectations for an assignment, providing focused feedback on works in progress, and grading final products (Andrade, 2000; Moskal, 2003; Popham, 1997). Although educators tend to define *rubric* in slightly different ways, a commonly accepted definition is a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor (see Appendix A for rubrics that fit this definition). Current books and articles on classroom assessment are rife with claims about the potential for student-involved assess- ment in general and rubrics in particular to increase stu- dents' self-efficacy and, as a result, lead to improvements in learning and achievement (e.g., Arter & McTighe, 2001; Quinlan, 2006; Stiggins, 2001). The assumption is that heightened self-efficacy is one of the mechanisms by which rubrics provide an advantage, yet no empirical evidence of a relation between rubric use and self-efficacy exists.

Some research suggests that rubric use can be related to improvements in the quality of students' writing and knowledge of the qualities of effective writing. Researchers of writers in Grades 3 and 4 (Andrade, Du, & Wang, in press) and in Grades 4, 5, and 6 (Ross, Rolheiser, & Hog-aboam-Gray, 1999) have shown a relation between writing scores and rubric-referenced student self-assessment. In a study of group learning in five/Grade 6 social studies class-es, Cohen, Lotan, Scarloss, Schultz, and Abram (2002) found that students who were informed of the evaluation criteria for written essays had higher quality discussions and better group products than did students who worked without knowing the criteria. Using path analysis, these authors concluded that knowledge of evaluative criteria had an indirect effect on essay scores, with group products and self-assessment (group discussions of the quality of their product) playing a key mediating role.

In a study of eighth-grade students' writing, Andrade (2001) showed that simply providing students with a rubric was associated with higher scores on only one of three essays; however, questionnaires administered at the end of the study revealed that students in the treatment group tended to identify more of the criteria by which their writing was evaluated. Andrade concluded that simply handing out and explaining a rubric can increase students' knowledge of the criteria for writing, but translating that knowledge into actual writing is more demanding. She recommended sustained attention to the process of assessing writing, including involving students in the design of rubrics by critiquing sample pieces of writing and by teaching students to self-assess their works in progress.

Andrade's (2001) recommendation regarding involving students in cocreating rubrics by critiquing examples is supported by research on the power of models in promoting skill acquisition. Zhu, Simon, and colleagues (Zhu, Lee, Simon, & Zhu, 1996; Zhu & Simon, 1987; Zhu, Zhu, Lee, & Simon, 2003) have demonstrated that studying worked-out examples of science or mathematics problems can help students acquire new information and skills, use the skills to solve new problems, and express solutions efficiently and

accurately. Wiggins (1998) argued that examples or models can be equally useful in teaching writing. Noting that the performance standards on rubrics are open to interpretation and that some students' views of "what it means to meet these criteria and the standard may be way off the mark" (p. 183), Wiggins recommended giving students models to promote more accurate analyses of the criteria in a rubric. Orsmond, Merry, and Callaghan (2004) agreed that a key factor in self-assessment is students' understanding of specific criteria and recommended the use of a subject-specific exemplar.

For these reasons, students in the treatment group in the present study were given a model essay or story and asked to generate a list of criteria for their writing assignments by listing the qualities that made the model effective. Because we needed to use similar or identical rubrics in different classes to make cross-class comparisons, students were not involved in co-creating entire rubrics. Rather, they were asked to generate a list of the criteria for their assignment, which invariably matched the rubrics that they were given during the next class.

Although the aforementioned research suggests that rubric use can promote academic achievement, there are no available studies that directly investigate the mechanism behind any rubric advantage. Many educators believe that student confidence or self-efficacy is behind the effect. *Self-efficacy* is an individual's belief in his or her capability to achieve a specific goal (Bandura, 2003). Pajares (2000) noted, "It's not just a matter of how capable you are, it's also a matter of how capable you think you are" (p. 13). He cited extensive research that has shown that students' self-efficacy exerts a powerful influence on their academic achievement, including writing (Pajares, 2003; Pajares & Johnson, 1996; Pajares & Valiante, 1997), even at the elementary and middle school levels (Pajares, Miller, & Johnson, 1999; Pajares & Valiante, 1999). Highly efficacious students tend to see difficult tasks as challenges to be met. Their efficacious outlook fosters intrinsic interest in activities and prompts them to work harder, persist longer, adopt what they believe are better strategies, and seek help from teachers and peers. In contrast, students with low self-efficacy tend to avoid challenging tasks and give up quickly (Bandura; Schunk, 2003).

Claims about the self-efficacy–boosting powers of rubrics are common, especially in books and articles written for teachers. For example, Arter and McTighe (2001) asserted that engaging students in generating and using criteria (a partial rubric) "increases student motivation, confidence, and achievement" (p. xi). Similarly, Ross (2006) argued that self-assessments that focus "student attention on particular aspects of their performance (e.g., the dimensions of the co-constructed rubric)" (p. 6) contribute to positive self-efficacy beliefs. Quinlan (2006) claimed that, when given a rubric, students "approach assignments with more confidence and resulting increased self-efficacy" (p. 119). Stix (1996) maintained that involving students in developing a rubric can

boost their confidence in their own abilities and their motivation to push past difficulties. Popular assessment expert Rick Stiggins (2001) concurred, claiming that "confidence is key to student success in all learning situations" (p. 43) and that involving students in the assessment process allows teachers to "tap an unlimited wellspring of motivation that resides within each learner" (p. 46).

In the present article, we admit to being skeptical of these claims, but there are reasons to believe them. Research has shown that there are a number of ways to boost self-efficacy, several of which may be enacted through a rubric. For example, in a study of undergraduate students' responses to rubric use over the course of a semester, Andrade and Du (2005) noted how students reported that having and using a rubric helped reduce anxiety about an assignment. Schunk (2001) noted that knowledge of specific performance standards can raise self-efficacy because progress toward an explicit goal is easy to gauge. Rubrics, by definition, provide specific performance standards and should boost self-efficacy.

There is some evidence that self-assessment or self-evaluation can also promote self-efficacy. For example, Paris and Paris (2001) reviewed research that suggests that self-assessment is likely to promote monitoring of progress, stimulate revision strategies, and promote feelings of self-efficacy. In a linear structural model, Wagner (1991, cited in Ross et al., 1999) found positive path coefficients from self-evaluation to self-efficacy. Schunk and Ertmer (1999) showed that "the opportunity for self-evaluation promoted self-efficacy" (p. 257). Schunk (2003) recommended giving students practice with criterion-referenced self-evaluation to develop and sustain self-efficacy for learning. Results from Kitsantas, Reiser, and Doster's (2004) study of 9th- and 10th-grade students who were learning to use presentation software showed that "among students who received organizational signals, those in the self-evaluation condition reported significantly higher levels of self-efficacy than did those in the no self-evaluation condition" (p. 284). In a qualitative study (Andrade & Du, 2005), undergraduates reported that criteria-referenced self-assessment made them feel more motivated and confident about their work.

We designed the present study to test the popular claims about the effects of rubric-referenced assessment, especially self-assessment, on elementary and middle school students' self-efficacy for a writing assignment. To investigate the influence of duration of exposure to rubrics, we examined short- and long-term rubric use. Because previous research has indicated that female students tend to have higher self-efficacy for writing than do boys (Pajares & Valiante, 1997, 1999), at least when researchers use measures that account for the tendency for girls and boys to use a different metric when providing confidence judgments (Pajares et al., 1999; Pajares & Valiante, 1999), in the present study, we also examined gender. The research questions that guided this study were the following:

Research Question 1: Is there a relation of short-term rubric use and elementary and middle school students' self-efficacy for a writing assignment?

Research Question 2: Is there a relation of long-term rubric use and self-efficacy for a writing assignment?

Research Question 3: Is there a gender difference in students' self-efficacy for a writing assignment?

Research Question 4: Does the effect of treatment differ by gender?

Answers to these questions will help researchers and teachers better understand and manage the relation between rubric use and students' confidence in and motivation for writing.

Method

Participants

Participants were 307 students in a convenience sample of volunteers in 18 elementary and middle school classes. After we accounted for missing data, including student absences from school during 1 or more days of the intervention, the actual sample size for the statistical analysis was 268. Of the classes, 9 were in a public school (School 1) with a population largely lower to middle class and White, 7 were in a private school for girls (School 2), and 2 were in a private school for boys (School 3). Both private school populations were largely middle to upper-middle class and White. All three schools were located in the Northeastern United States. Of the 18 classes, 13 were English or language arts, and 5 were history or social studies.

Table 1 presents the demographic information for the participants. Of the participants, 167 (62%) attended the public school, and the remaining 101 (38%) attended one of the two private schools. Among the participants, 99 (37%) were boys, and 169 (63%) were girls.

TABLE 1. Demographic Information for Participants (*N* = 268)

Demographic	School 1	School 2	School 3	Total
Condition				
Comparison	77	43	17	137
Treatment	90	30	11	131
Gender				
Female	96	73	0	169
Male	71	0	28	99
Grade level				
Grades 3–4	122	8	17	147
Grades 5–7	45	65	11	121
Ethnicity				
White	160	62	20	242
African American	6	2	1	9
Other	1	9	6	17
Total	167	73	28	268

Participating students' grade levels ranged from Grade 3 to Grade 7. The sample comprised 54 third-grade students (20.1%), 93 fourth-grade students (34.7%), 41 fifth-grade students (15.3%), 56 sixth-grade students (20.9%), and 24 seventh-grade students (9.0%). Ethnicity information was available for 239 participants in the study. The majority of those participants (*n* = 242; 90.3%) were White. Special needs information was available for the participants from the public school (School 1) and the private school for girls (School 2). Five student participants were identified by their school as having special needs in reading, 6 as having special needs in reading and writing, and 2 as having English as a second language.

The treatment and comparison groups consisted of intact classes, nine in each condition. We made assignments to the treatment or comparison group systematically, in terms of two variables: (a) the degree to which the classroom teachers had already used rubrics with the participating classes and (b) grade level. We took this approach to balance prior experience with rubrics and to ensure comparable numbers of students in each grade. As shown in Table 2, the treatment group consisted of four classes in Grades 3–4 and five classes in Grades 5–7; the comparison group comprised five classes in Grades 3–4 and four classes in Grades 5–7. Of the nine treatment classes, five had not used rubrics and four had used rubrics at least once or twice. Of the nine comparison classes, four had not used rubrics and five had used rubrics at least once or twice.

Instruments

Self-efficacy. Self-efficacy was measured through an adapted version of the Writing Self-Efficacy Scale used by Pajares, Hartley, and Valiante (2001). The 11-item writing self-efficacy scale (see Appendix B) measures individuals' confidence in their writing abilities, including their skill in handling commonly assessed qualities of writing: ideas and content, organization, paragraph formatting, voice and tone, word choice, sentence fluency and conventions (e.g., the 6+1 Trait Writing Method; see Culham, 2003; Spandel & Stiggins, 1997). Students were instructed to rate their confidence levels on a scale of 0–100. The 0–100 format was selected over the traditional Likert-type scale because Pajares et al. documented that a scale with a 0–100 format was psychometrically stronger than a 1–10 scale in regard to factor structure and internal consistency. Pajares et al. also found that, compared with the traditional Likert-type scale, the 0–100 scale has better discrimination and stronger relations with various achievement indexes. For the sample in the present study, the measure yielded alpha reliabilities of .91, .92, and .91 for the three administrations of the self-efficacy instrument, respectively.

Previous exposure to rubrics. Students' exposure to rubrics was measured in two ways. Teachers who volunteered to participate in the study were asked about their rubric use with the class or classes involved in the research. Their

TABLE 2. Number of Classes Assigned to Treatment and Comparison Conditions, by Grade Level and Teachers' Prior Rubric Use

| Group | Teacher's prior rubric use (*Yes*) | | Teachers' prior rubric use (*No*) | | Total |
	Grades 3–4	Grades 5–7	Grades 3–4	Grades 5–7	
Treatment	1	3	3	2	9
Comparison	2	3	3	1	9

responses were confirmed by observation in their classroom and categorized according to the 0–4 scale in Table 3. Of the 15 teachers in the study, 3 were categorized as Level 0, 3 as Level 1, 5 as Level 2, 3 as Level 3, and 1 as Level 4. The teacher rating of prior rubric use by the treatment group was not statistically different from that of the comparison group, $t(16) = 0.80$, $p = .44$. This variable was used to assign classes to treatment condition and was not included in the analysis because it is a class-level variable.

Data for a student-level variable regarding prior rubric use were generated by asking students to answer two questions on a questionnaire administered at the beginning of the study: (a) "Has your teacher for *this class* ever given you a rubric for a writing assignment? (*Yes* or *No*)" and (b) "If yes, about how many times has your teacher given you a rubric for a writing assignment? (1–2 times, 3–5 times, 6–10 times, 10 or more times)." Class averages of students' responses ranged from 0 (*No, my teacher has not given a rubric for a writing assignment*) to 3.28 (*Yes, my teacher has given a rubric for a writing assignment 3–5 times*). The average rating for the treatment group was 1.41 ($SD = 1.44$). The average rating for the comparison group was 1.05 ($SD = 0.97$). The average student rating of previous rubric use in the treatment group was higher than that of the comparison group, $t(266) = 2.46$, $p = .015$.

The data collected from the questionnaires were used as a measure of long-term rubric use. Data were collected between January and March of 2006. Because each class began meeting in September, we defined *long-term rubric use* as use for between 5 and 7 months.

Writing assignments. Each class was asked to do a writing assignment. Of the 18 classes, 2 third-grade classes (1 treatment and 1 comparison) wrote stories. The remaining 16 classes (8 treatment and 8 comparison), including 2 more third-grade classes, wrote persuasive essays. The writing process in each class resembled a writers' workshop: Students engaged in some form of prewriting, wrote rough drafts, received feedback from the classroom teacher, and wrote final drafts.

TABLE 3. Teachers' Reported Level of Prior Rubric Use with Participating Classes

Level	Description of Rubric Use
0	I do not use rubrics in this class.
1	I use rubrics in this class. I create the rubric and discuss the expectations with students but do not hand out the rubric before students begin an assignment. I use the rubric to grade student work.
2	I use rubrics in this class. I create the rubric and hand out and review the rubric with students before they begin to work on their assignments. I use the rubric to grade student work.
3	I use rubrics in this class. I create and review the rubric with students before they begin their assignments. I ask students to use the rubric to evaluate their own or others' writing some of the time. I use the rubric to grade student work.
4	I use rubrics in this class. I sometimes or always create the rubric with my students. We review the rubric before they begin their assignments. I ask students to use the rubric to evaluate their own and their peers' work most of the time. I use the rubric to grade student work.

Procedures

Table 4 summarizes the sequence of events followed by each class. To ensure the fidelity of the treatment, the first author co-led Class Periods 1, 2, and 4 with the classroom teachers.

The treatment condition differed from the comparison condition in three ways: The students in the treatment group (a) read a model story or essay, discussed its strengths and weaknesses, and generated a list of qualities of an effective story or essay; (b) received a written rubric (Appendix A); and (c) used the rubric to self-assess their first drafts. The students in the comparison group did not read a model but did generate a list of qualities of an effective story or essay. The comparison group did not receive a rubric. Students in the comparison group were asked to review their first drafts and note possibilities for improvement in the final draft. They did not self-assess their drafts according to a rubric.

Models and criteria generation. The treatment group was

TABLE 4. Sequence of Events, by Condition and Class Period

Group	Class Period 1	Class Period 2	Class Period 3	Class Period 4	Class Period 5+
Treatment	1. Introduce assignment. 2. Read and discuss model story or essay. 3. Generate list of qualities of an effective story or essay. 4. Practice self-efficacy rating. 5. Administer first self-efficacy assessment.	1. Hand out and discuss rubric 2. Administer second self-efficacy assessment 3. Students do prewriting (e.g., outlining, brainstorms)	Students write first drafts.	1. Students use rubric to self-assess first drafts. 2. Administer third self-efficacy assessment.	1. Classroom teacher gives each student feedback. 2. Students write final drafts.
Comparison	1. Introduce assignment. 2. Generate list of qualities of an effective story or essay. 3. Practice self-efficacy rating. 4. Administer first self-efficacy assessment.	1. Administer second self-efficacy assessment. 2. Prewriting (e.g., outlining, brainstorms).	Students write first drafts.	1. Students self-assess drafts without rubric. 2. Administer third self-efficacy assessment.	1. Classroom teacher gives each student feedback. 2. Students write final drafts.

qualities of effective writing and scaffold thinking about the criteria for students' own essays or stories. Researchers have argued that the process of generating criteria is beneficial to students (Andrade, 2000, 2001; Ross et al., 1999). However, for research purposes, the rubrics given to different classes in the treatment group were the same; different classes did not cocreate idiosyncratic rubrics.

Self-assessment. The rubric-referenced self-assessment done by students in the treatment group was guided by the first author. Students were asked to underline key phrases in the rubric with colored pencils (e.g., "clearly states an opinion") and then underline or circle in their drafts the evidence of having met the standard articulated by the phrase (e.g., his or her opinion). If they found they had not met the standard, they were asked to write themselves a reminder to make improvements when they wrote their final drafts.

Self-efficacy ratings. All student participants were administered the Writing Self-Efficacy Scale (see Appendix B) three times: (a) during Class Period 1, after the writing assignment was introduced; (b) during Class Period 2, after the rubric was handed out (treatment) or not handed out (comparison); and (c) during Class Period 4, after the rubric-referenced self-assessment of drafts (treatment) or review of drafts (comparison). On the advice of Bandura (2006), the first administration of the instrument was pre-

were asked to rate their confidence that they could jump increasing distances (three, five, and seven floor tiles) on a scale of 0–100 and to then actually attempt the jumps.

In-class writing. Students were given class time to complete each step of the writing process. The amount of class time devoted to writing (not instruction or treatment) varied by class, from 90 to 265 min. The amount of time devoted to writing was determined by the teachers, who were encouraged to conduct their lessons as they typically did. No upper or lower limit on writing time was set by the researchers. The average time spent on writing by the treatment group was 159 min (SD = 52 min). The average time spent on writing by the comparison group was 156 min (SD = 24 min). On average, the treatment and comparison groups had equivalent amounts of class time for writing, $t(16)$ = 0.20, p = .84. Writing time was not significantly correlated with any of the three self-efficacy ratings (r = .28, p = .26; r = .25, p = .32; and r = .11, p = .65; respectively).

Results

Preliminary Analysis

For the full sample (N = 268), the mean of the self-efficacy rating was 82.6 (SD = 17.10) for the first adminis-

and 87.7 (SD = 14.16) for the third administration. The mean rating on the first self-efficacy scale is comparable to those of previous research with elementary and middle school students' mean prewriting self-efficacy ratings of 80, 84, 83, and 78 (SDs = 14.6, 12.6, 13, and 17.4, respectively; Pajares et al., 1999; Pajares & Valiante, 1999). The means and standard deviations of self-efficacy ratings at each of the three times are compared by gender in Table 5.

Differences by condition. The means and standard deviations of self-efficacy ratings at the three times by group (treatment or comparison) are also shown in Table 5. A t test analysis showed no difference between the treatment and comparison groups in scores for the first two administrations of the self-efficacy instrument. At Time 3, the average writing self-efficacy score of the treatment group was higher than that of the comparison group. The difference approached but

did not reach statistical significance, $t(266)$ = 1.79, p = .075. Figure 1 shows the pattern of change in self-efficacy for the treatment and comparison groups.

Differences by gender. A t test analysis showed gender differences in self-efficacy scores, favoring girls for the first administration of the self-efficacy instrument, $t(266)$ = 2.48, p < .05. The differences between boys and girls approached significance for the third administration of the instrument, $t(266)$ = 1.92, p = .056. However, at Time 2, the differences in the average writing self-efficacy scores for girls and boys were not statistically significant. Figure 2 shows the pattern of change in self-efficacy for boys and girls.

Differences by grade level and school type. A t test showed no statistically significant differences in self-efficacy ratings across grade levels (Grades 3–4 vs. Grades 5–7) for any of the three administrations of the self-efficacy assessment:

TABLE 5. Means and Standard Deviations of Self-Efficacy Scores at the Three Different Times, by Condition and Gender (N =268)

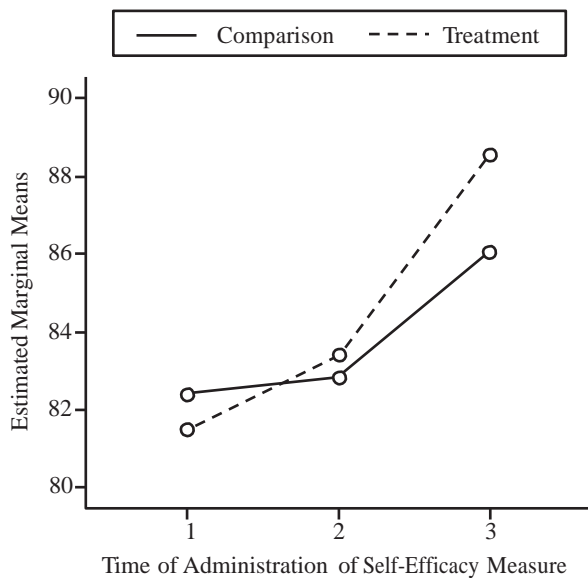| Variable	Time 1					Time 2					Time 3				
	M	SD	t	p	ES	M	SD	t	p	ES	M	SD	t	p	ES
Condition			0.14	.885	.02			−0.49	.626	.06			−1.79	.075	.22
Comparison	82.76	16.032				82.94	17.764				86.20	15.016			
Treatment	82.45	18.204				83.98	16.905				89.29	13.078			
Gender			2.48	.014	.31			1.32	.189	.17			1.92	.056	.24
Female	84.57	15.712				84.51	17.058				88.97	13.532			
Male	79.25	18.845				81.63	17.709				85.56	14.996			

Note. ES = effect size.

FIGURE 1. Estimated marginal means of writing self-efficacy scores across three time points, by treatment and comparison condition (N = 268).

FIGURE 2. Estimated marginal means of writing self-efficacy scores across three time points, by gender (N = 268).