Howell, R. J. (2011). Exploring the impact of grading rubrics on academic performance: Findings from a quasi-experimental, pre-post evaluation. *Journal on Excellence in College Teaching*, 22 (2), 31-49.

Exploring the Impact of Grading Rubrics on Academic Performance: Findings From a Quasi-Experimental, Pre-Post Evaluation

Rebecca J. Howell
University of Alabama

*This purpose of this pre-post, quasi-experimental evaluation was to explore the impact of grading rubric use on student academic performance. Cross-sectional data were derived from 80 undergraduates enrolled in an elective course at a research university during spring and fall 2009. The control group (*n = *41), who completed the course's Assignment #2 without a grading rubric, scored significantly lower, on average, than the treatment group (*n = *39), who completed the same assignment, but with access to a grading rubric. The grading rubric constituted an important predictor of assignment performance, the magnitude of which was stronger than college year, major, pre-test score, and gender. Suggestions are provided for future research.*

Beginning in the mid-1980s, higher education began to shift from an emphasis on the traditional paradigm of testing knowledge and teacher-centered learning to a paradigm characterized by active, student-centered learning and thoughtful, deliberate assessment (Cox & Richlin, 1993; Wiggins, 1998). Rather than simply regurgitating factual information, the modern-day student is expected to work toward solving open-ended "real world" problems and tasks by actively demonstrating higher-order cognitive competencies, such as critical and reflective thinking (Brown, Bull, & Pendlebury, 1997; Wiggins, 1998).

It was within this re-direction in pedagogy, and in partial response to the accountability pressure exerted by multiple parties (the public, state and federal governments, and regional accrediting associations), that

objective and evidence-based assessment was born (Dochy, Gijbels, & Segers, 2006). This empirical approach to learning assessment remains a permanent fixture of contemporary academia. Not only has there has been federal talk of national standardized testing in higher education (Arenson, 2006), but some state legislatures have enacted assessment-based performance standards for institutions in making budget allocation decisions (Wellman, 2001), and each of the six regional accrediting associations have established assessment-related standards of effectiveness. Further, it has been noted that in recent years, assessment constitutes the area in which institutions have been most likely to be deemed as non-compliant by the Southern Association of Colleges and Schools (Weiss, Cosbey, Habel, Hanson, & Larsen, 2002).

Course instructors employ a variety of strategies to enhance student learning. At the foundational level, courses are being designed with specific learning outcomes in mind, and more and more instructors are linking these outcomes to specific course content and methods of evaluation (Dochy et al., 2006; Peat, 2006; Peat & Moriarty, 2009). Coincident with this initiative is the commonly used mode of assessment that involves evaluating open-ended academic work, such as reflective essays, case studies, and papers, with the guidance of grading (or scoring) rubrics.

In its most basic form, a grading rubric is a matrix that provides levels of achievement for a set of criteria or dimensions of quality for a given type of performance (Allen & Tanner, 2006). Rubrics serve multiple functions. They formulate standards for achievement, provide an objective way to grade work and provide students with appropriate feedback, and may make expectations clearer to students, particularly when a rubric is used as a guide while completing the task at hand (Allen & Tanner, 2006; Arter & McTighe, 2001; Perlman, 2003). Although rubrics differ in terms of length and content, they tend to assess either a specific task or a more general type of performance. In addition, rubrics generally are either analytic or holistic in nature (Jonsson & Svingby, 2007). Within each level of achievement, analytic rubrics assess multiple aspects of performance. In contrast, holistic rubrics are designed to assess the overall quality of the work in question.

Extant Research

A relatively large literature base exists on particular aspects of rubric use, with much attention paid to the elements of design, development, and implementation; the philosophical rationale of use; opinions about the consequent benefits of use; interrater reliability; and construct validity

(see, for example, Hafner and Hafner, 2003; Jonsson and Svingby, 2007; Stemmack, Konheim-Kalkstein, Manor, Massey, and Schmitz, 2009; Thaler, Kazemi, and Huscher, 2009). With respect to academic performance benefits, the general consensus among instructors and higher administration, alike, is that learning and academic performance is positively enhanced when rubrics are used to evaluate student performance (Dahm, Newell, & Newell, 2004; Jonsson & Svingby, 2007; Moriarty, 2006; Peat, 2006; Peat & Moriarty, 2009). A limited number of studies indicate that students and teachers perceive that rubrics clarify expectations and make the grading standards more explicit (Bissell & Lemons, 2006; Dahm et al., 2004; Frederiksen & Collins, 1989). There is, however, a paucity of literature on the actual effectiveness of rubric use on students' academic performance.

Jonsson and Svingby's (2007) qualitative meta-analysis constitutes one of the most comprehensive reviews conducted on the rubrics literature. Their findings indicate that only 10 of the 75 studies on grading rubrics published within the past 40 years investigated the impact of their use on student improvement; the bulk of these studies did not investigate the outcome of academic performance. Unfortunately, the mixed findings of this small body of research precluded the ability to "draw any conclusions about student improvement related to the use of rubrics from this material" (Jonsson & Svingby, 2007, p. 139). One study (Andrade, 1999), which evaluated the usefulness of a scoring rubric in the self-assessment of junior high school students, found no significant difference, while another study (Toth, Suthers, & Lesgold, 2002) found that rubric use can improve student performance in the hard sciences, but only when additional instructional methods are employed. There is some indication that rubrics can elicit general improvements in writing (Brown, Glasswell, & Harland, 2004) and the understanding of science among middle-school students (Mullen, 2003). The results from the remaining studies also suggest that rubric use constitutes a promising endeavor, as specific areas of performance may be enhanced, such as writing development among deaf students (Schirmer, Bailey, & Fitzgerald, 1999) or critical thinking, problem-solving, and self-assessment among students from the general population (Green & Bowser, 2006; Sadler & Good, 2006; Schafer, Swanson, Bene, & Newberry, 2001; Schamber & Mahoney, 2006).

Given the lack of quantitative research on this issue, this study sought to explore whether grading rubric use enhances the academic performance of college students. It tested the following hypothesis: *Grading rubric use exerts an important, positive impact on students' academic performance*.

Method

Subjects and Setting

A quasi-experimental, pre-post evaluation was carried out in a 200-level, undergraduate elective course, Juvenile Delinquency, that was offered during spring and fall 2009 at a large research university. The researcher (the author) served as the instructor of the course. As a campus-wide elective, the course is accepted by the university to satisfy a requirement of the general education curriculum. Sociology minors and criminal justice majors and minors also can complete the course and apply the elective hours toward their respective programs of study.

The university in which the course was taught is located in a large Southeastern U.S. town. As the seat of county government, the community's economic base is primarily industrial, manufacturing, and service in nature. The university itself serves as the community's largest employer, while the town in which it is located serves as a "bedroom community" for a major city within a relatively short commuting distance.

Assignment #2 for Juvenile Delinquency, worth a total of 10 points (out of 500 points in the course), constituted the test instrument. The purpose of the assignment was to have students actively apply a delinquency theory to a "real world" delinquent act, in this case, shoplifting. Students were given the assignment directions in Figure 1.

During spring 2009, 41 students enrolled in the course completed Assignment #2 by following the assignment directives presented above. This version of Assignment #2 did not contain a grading rubric; students simply were informed that the assignment was worth 10 points, and they were asked to read and follow the assignment directions.

A holistic, task-specific grading rubric (see Table 1) was developed during summer 2009 and added to the end of the assignment. This rubric contained four levels of student performance: beginning, developing, accomplished, and exemplary. Achievement criteria were detailed for each respective level along with the corresponding number of points. The inclusion of this grading rubric was the only change made to the assignment. During fall 2009, Assignment #2, which now contained the grading rubric, was distributed and completed by 39 students.

Fifty-three and 51 students were enrolled in the course during spring and fall 2009, respectively. The Assignment #2 completion rate for students in the spring 2009 course was 77%, while 76% of students enrolled during fall 2009 completed the assignment. The stringent attendance-assignment policy, which was in force during both semesters, may explain, in part, why the assignment completion rates were not higher. In an effort to re-

Figure 1
Juvenile Delinquency Assignment #2

Assignment #2:

Use your own words (no quotations; paraphrasing with citations is permitted), and respond in well-written, complete, and typed sentences.

Step #1: Discuss what Routine Activities Theory is about, including the theory's assumptions of human nature and behavior and the elements that the theory comprises.

Step #2: After reading the case study (see below), use Routine Activities Theory to explain Jen's shoplifting behavior. In your discussion.

Be sure to include details from the case study to illustrate each element of the theory. Be sure your explanation is complete and specific.

enforce class attendance, the attendance-assignment policy was such that if students wanted the instructor to grade their respective assignments, students had to have attended class on the day a given assignment was distributed and due.

Procedure

Students from both semesters received the same instructional materials, were taught by the same instructor using the same course content, were taught in the same manner (that is, PowerPoint lectures, class discussion, videos, and optional chapter learning objectives), received the same number of assignments and exams, and submitted required work that the instructor personally graded. Each class also met on the same days of the week (Tuesdays and Thursdays) and during the same class period (8:00-9:15 a.m.) across a 15-week semester.

No students from either class were aware that the assignment would be or had been redesigned, and no students enrolled in the course during the spring took the course again in the fall. Students from both semesters completed the assignment during the same time period (the fourth week) and the instructor graded all assignments from both semesters. The research was approved by the university's Institutional Review Board for the Protection of Human Subjects.

Table 1
Grading Rubric

Performance Level	*Criteria*	*Comments*	*Point Breakdown*
Exemplary	• Well-written and clear • Accurate theoretical application that discusses the theory, identifies all theoretical elements, and explains (in full) Jen's delinquent behavior by linking the corresponding theoretical elements to case study information/details		10 pts.
Accomplished	• Relatively well-written and/or fairly clear • Accurate theoretical application that discusses the theory, identifies all theoretical element, and explains (in full) Jen's delinquent behavior by linking the corresponding theoretical elements to case study information/details		9 pts.

Developing	• Moderately well-written and/or somewhat clear • Somewhat accurate theoretical application that discusses the theory (in part), identifies the bulk of the theoretical elements, and explains (in part) Jen's delinquent behavior by linking the corresponding theoretical elements to case study information/details	7-8 pts.
Beginning	• Poorly written (major spelling, grammar issues) and/or unclear • Inaccurate theoretical application that discusses the theory (in part), identifies some of the theoretical elements, and explains (in part) Jen's delinquent behavior by linking the corresponding theoretical elements to case study information/details	1-6 pts.

Measures

Data were derived from four sources: Assignment #2 scores, the course pre-test scores, student information sheets, and detailed class listings. A total of six variables were examined. The dependent variable, *Assignment #2 Score*, constituted a continuous measure and indicated how well a given student performed on Assignment #2 of the course. Measured on a 100-point scale, "0" was indicative of 0%, while "100" was indicative of a perfect score, or 100%.

Treatment, a dichotomous, independent variable, constituted the independent variable of interest. The control group consisted of those students who were enrolled in the spring 2009 semester and who completed Assignment #2 without the aid of the grading rubric. The treatment group consisted of students from the fall 2009 course who completed Assignment #2 with the use of the grading rubric.

Some empirical evidence suggests that academic performance may be impacted positively by higher year in college (Cejda, Kaylor, & Rewey, 1998), higher baseline course-specific knowledge (Halloun & Hestenes, 1985), and gender, with females slightly more likely to perform better academically than males (Hyde, Fennema, & Lamon, 1990; Pomerantz, Altermatt, & Saxon, 2002). In light of these associations, the effects of four possible explanatory factors were controlled: *College Year*, *CJ Major*, *Pre-Test Score*, and *Gender*. Data on *College Year* (freshman, sophomore, junior, senior) and *Gender* were collected via an information sheet that students voluntarily completed on the first day of class during both semesters. Data on *CJ Major*, a dichotomous variable (non-criminal justice major versus criminal justice major), were obtained from the detail class listing for each semester. Measured on a 100-point scale, *Pre-Test Score* constituted students' scores on the pre-test instrument for the course.

Each semester, the instructor administered the same pre-test instrument to students on the first day of class. The purpose of this instrument was to assess, at baseline, what course-specific knowledge students had upon beginning the class. This instrument was completed on a voluntary basis, and scores did not count toward students' final course grades.

Results

Two major types of analyses were conducted: independent samples *t* tests and ordinary least squares regression (OLS). In determining whether the treatment and control groups differed significantly, on average, with respect to the dependent variable, *Assignment #2 Score*, a *t* test for inde-

pendent samples was calculated. This statistical technique provided a simple test of the differences in the means scores on Assignment #2 for the control and treatment groups. Results indicated that the treatment group's mean (83.84%) on Assignment #2 was significantly higher than the control group's mean (72.19%) for the assignment (t = -4.267; df = 78; p < .01). To test the hypothesis that the grading rubric enhanced performance on Assignment #2, I turned my attention to developing and analyzing the multivariate regression model.

To prepare the data for multivariate analysis, the control and treatment groups were collapsed into one total sample (N = 80). Descriptive statistics then were calculated for each variable, bivariate correlations between each of the measured were assessed, independent samples t tests were conducted for each covariate to assess group heterogeneity, and then the OLS model was developed.

Table 2 details the descriptive statistics for all of the variables. Across both semesters, scores on Assignment #2 ranged from a low of 40% to a high of 100%, with an average of 77.87%. Less than half (49%) of the sample completed Assignment #2 with the aid of the grading rubric. Finally, across both courses, on average, students were female, in their junior year of college, enrolled in a non-criminal justice program of study, and scored 53.51% on the course pre-test.

Bivariate correlations (see Table 3) were evaluated in an effort to identify problematic collinearity between any two measures. Bivariate collinearity was not a problem; all coefficients were smaller than .500 (George & Mallery, 2001). Two of the four significant correlations centered on Assignment #2. As compared to their counterparts, the students who performed better on Assignment #2 tended to be those who were provided the grading rubric (r = .435; p < .05) and those who did well on the pre-test (r = .277; p < .10). Students from the control group tended to be juniors and seniors as opposed to freshmen and sophomores (r = -.329; p < .05). Across both semesters, slightly more males than females were majoring in criminal justice (r = .231; p < .10).

Given the lack of random assignment to treatment and control groups, it was important to assess the degree of group heterogeneity, because this type of sampling bias is a threat to the internal validity of any regression findings (Shadish, Cook, & Campbell, 2001). In calculating four independent-samples t tests (see Table 4), the treatment and control group's mean values were compared on each of the four covariates (*College Year*, *CJ Major*, *Pre-Test Score*, and *Gender*) that were subsequently entered into the OLS model. In general, the results indicated that the groups were relatively homogenous with respect to the four covariates. Students in the treatment

Table 2
Variable Descriptives (*N* = 80)

Variable	*Min*	*Max*	*Mean*	SD
Assignment #2 score	40.000	100.000	77.875	13.472
Treatment	0.000	1.000	0.487	0.503
College year	1.000	4.000	2.920	0.924
CJ major	0.000	1.000	0.450	0.500
Pre-test score	20.000	82.000	53.512	11.423
Gender	0.000	1.000	0.325	0.471

Table 3
Bivariate Correlations (*N* = 80)

Variable	*Assignment #2 score*	*Treatment*	*College Year*	*CJ major*	*Pre-Test Score*	*Gender*
Assignment #2 score	---					
Treatment	.435**	---				
College year	.109	-.329**	---			
CJ major	.162	.073	-.172	---		
Pre-test score	.277*	.134	.126	.138	---	
Gender	-.149	.017	-.118	.231*	.049	---

Note. *$p < .10$; **$p < .05$

Table 4
Independent-Samples *t*-Test Results (*N* = 80)

	Levene's Test for Equality of Variances		t *Test for Equality of Means*		
Covariate	***F***	**Sig.**	***t***	***df***	**Sig.**
College Year	7.322	.008	3.048	68.464	.003
CJ Major	1.067	.305	-.645	78.000	.521
Pre-Test Score	.347	.557	-1.198	78.000	.235
Gender	.094	.760	-.153	78.000	.879

and control groups were similar with respect to *CJ Major*, *Pre-Test*, and *Gender*. Group heterogeneity on *College Year* was observed, however, with the treatment group containing substantively less upper-class students than the control group ($t = 3.048$; $df = 68.464$; $p < .05$).

The results for the OLS model are presented in Table 5. Together, the variables explain 34.3% of the variance in *Assignment #2 Score*. Two significant predictors were identified. Most important, use of the grading rubric, as measured by *Treatment*, exerted an independent, positive, and significant impact on *Assignment #2 Score* ($ß = .488$; $p < .01$). In particular, regardless of year in college, college major, pre-test score, and gender, students from the treatment group performed better on Assignment #2, on average, than students from the control group. Furthermore, a comparison of the size of the standardized regression coefficients (Beta, or *ß*) indicates that *Treatment* constituted the strongest predictor in the model; its effects were nearly double that of those posed by *College Year* ($ß = .261$; $p < .05$).

The only other significant predictor of performance on Assignment #2 was *College Year* ($r = .261$; $p < .05$), which also was positively associated with the assignment scores. The remaining variables did not make significant contributions; however, the directional effects observed are consistent with those yielded by the research. Specifically, criminal justice majors ($r = .191$; $p < .058$), students who scored relatively high on the pre-test instrument ($r = .106$; $p < .106$), and females ($r = -.179$; $p < .070$) tended to score higher on Assignment #2 than their respective counterparts.

Limitations and Suggestions for Future Research

Suggestions for future research stem largely from the study's two major limitations. The first limitation involves intra-rater reliability. There are three major explanations for variability in academic performance between a control and treatment group: actual group differences in performance, group differences in the grader's judgment, and group differences in expected tasks (Brown et al., 1997; Shavelson, Baxter, & Pine, 1991). The latter issue is not a likely alternative explanation for the observed association between *Treatment* and performance on Assignment #2, because the same assignment and directions were distributed to both the treatment and control group; the only difference was that the treatment group's assignment contained the grading rubric.

In contrast, group differences in grader judgment (that is, grading bias) remain a plausible alternative explanation for the positive rubric-performance association; it was beyond the scope of the study to rule out this validity threat. Although this limitation poses some concern for the

Table 5
OLS Results (*N* = 80)

Predictor	b	SE	ß
Treatment	13.081	2.722	.488**
College Year	3.804	1.511	.261*
CJ Major	5.131	2.668	.191
Pre-Test Score	.190	.116	.161
Gender	-5.104	2.780	-.179
$R^2 = .343 < .01$			

Note. *b* = unstandardized coefficient; SE = standard error; *ß* = Beta, or standardized coefficient.
*$p < .05$; **$p < .01$

validity of the rubric-performance association that was observed, it is important to point out that the impact of this threat on the findings likely is minimal. Seven research projects reviewed by Jonsson and Svingby (2007) investigated the intra-rater reliability of rubric scoring. Nearly all of these studies reported Cronbach's alpha values that exceeded the commonly accepted value of .70. Thus, while intra-rater reliability within this study likely was not exceedingly low, this threat to the internal validity of the results remains plausible because it was not objectively ruled out. Future research in this area should assess the impact of this threat on the validity of findings.

A second limitation is that the study did not employ a true experimental research design, and more than half of the variation in the scores on Assignment #2 was not explained by the constructs that were investigated. Because it is probable that factors other than those investigated made a contribution to assignment performance, the findings should be interpreted with some caution. One such explanatory factor may be grade point average (GPA), as research indicates that a linear relationship tends to exist between overall academic performance and performance at the course level (Kleijn, Ploeg, & Topman, 1994; Potolsky, Cohen, & Saylor, 2003; Sansgiry, Bhosle, & Sail, 2006; Suda, Franks, McKibbin, Wang, & Smith, 2008). I attempted to control for the effects of GPA in the study, but it was not feasible given the difficulty of tracking down and securing informed consent from the students in both the control and treatment

groups. A substantial number of these former students have graduated from the institution since taking the course. Not only can further research attempt to replicate the results reported here, but it may be worthwhile to consider the effects of GPA on any course-specific performance variables that are investigated.

In a similar vein, given the inability to assign students randomly to the treatment and control groups, sampling bias remains a plausible threat to the internal validity of the results. An effort was made to gauge, in part, the degree to which group heterogeneity existed by conducting independent-samples *t* tests for each of the four covariates that were subsequently entered into the OLS model. Overall, the findings were favorable, suggesting that no real group differences existed with respect to *CJ Major*, *Pre-Test*, and *Gender*. Although the control group contained substantively more upper-class students than freshmen and sophomores, the significant, positive effects posed by *College Year* in the OLS model indicated that this group difference in *College Year* served, if anything, to attenuate the relationship between the treatment and Assignment #2, not to inflate this significant association. Again, however, it is possible that the groups may differ with respect to variables that were not measured in this research (for instance, GPA).

Conclusions

Student assessment remains an important issue in academia. One way to evaluate academic performance at the course level is to grade work using a standardized rubric. The use of grading rubrics among faculty is widely cited in the literature (see Jonsson and Svingby, 2007), but there is little quantitative research on this tool's value for enhancing academic performance. The multivariate, quasi-experimental findings reported here favor such use. Providing students with an assignment-embedded grading rubric played a substantive role in positively impacting assignment performance, regardless of students' year in college, major program of study, baseline course-specific knowledge, and gender.

The favorable light that this research casts on the continued and increased use of rubrics may prompt some to ask: What exactly is it about rubric use that contributes to solid academic performance? Tentatively, the limited extant research (Bissell & Lemons, 2006; Dahm et al., 2004; Frederiksen & Collins, 1989) suggests that students may not simply write to the rubric, but are provided clearer expectations than when no rubric is given. Of benefit to higher education pedagogy is additional research in this area that supplements quantitative data on rubric impact with

qualitative data about why rubrics "work" and derived from students who use this pedagogical tool. Students also likely will benefit if faculty who use rubrics actively encourage these learners to take advantage of the expectations detailed within.

References

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *Life Sciences Education*, *5*, 197-203.

Andrade, H. G. (1999, April). *The role of instructional rubrics and self-assessment in learning to write: A smorgasbord of findings.* Paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Arensen, K. W. (2006, February 9). Panel explores standard tests for colleges. *The New York Times*. Retrieved from http://www.nytimes.com

Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom*. Thousand Oaks, CA: Corwin Press.

Bissell, A. N., & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *Bioscience*, *56*, 66-72.

Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.

Brown, G. T., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, *9*, 105-121.

Cejda, B. D., Kaylor, A. J., & Rewey, K. L. (1998). Transfer shock in an academic discipline: The relationship between students' major and their academic performance. *Community College Review*, *26* (3), 1-13.

Cox, M. D., & Richlin, L. (1993). Emerging trends in college teaching for the 21st century: A message from the editors. *Journal on Excellence in College Teaching*, *4*, 1-7.

Dahm, K. D., Newell, J. A., & Newell, H. L. (2004). Rubric development for assessment of undergraduate research: Evaluating multidisciplinary team projects. *Chemical Engineering Education*, *38* (1), 68-73.

Dochy, F., Gijbels, D., & Segers, M. (2006). Learning and the emerging new assessment culture. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present, and future trends* (pp. 191-206). Oxford, Amsterdam: Elsevier.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*, 27-32.

George, D., & Mallery, P. (2001). *SPSS® for Windows® step by step: A simple*

guide and reference (3rd ed.). Boston: Allyn and Bacon.
Green, R., & Bowser, M. (2006). Observations from the field: Sharing a literature review rubric. *Journal of Library Administration*, *45*, 185-202.
Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, *25*, 1509-1528.
Halloun, I., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, *53*, 1043-1055.
Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107* (2), 139-155.
Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*, 130-144.
Kleijn, W., Ploeg, H., & Topman, R. (1994). Cognition, study habits, test anxiety, and academic performance. *Psychological Reports*, *75*, 1219-1226.
Moriarty, L. J. (2006). Investing in quality: The current state of assessment in criminal justice programs. *Justice Quarterly*, *20* (4), 409-427.
Mullen, Y. K. (2003). *Student improvement in middle school science*. Unpublished master's thesis, University of Wisconsin, Madison, WI.
Peat, B. (2006). Integrating writing and research skills: Development and testing of a rubric to measure student outcomes. *Journal of Public Affairs Education*, *12* (3), 295-311.
Peat, B., & Moriarty, L.J. (2009). *Assessing criminal justice/criminology education: A resource handbook for educators and administrators*. Durham, NC: Carolina Academic Press.
Perlman, C.L. (2003). Performance assessment: Designing appropriate performance tasks and scoring rubrics. In C. Boston (Ed.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 497-506). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology*, *94* (2), 396-404.
Potolsky, A., Cohen, J., & Saylor, C. (2003). Academic performance of nursing students: Do prerequisite grades and tutoring make a difference? *Nursing Education Perspectives*, *24* (5), 246-250.
Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*, 1-31.
Sansgiry, S., Bhosle, M., & Sail, K. (2006). Factors that affect academic

performance among pharmacy students. *American Journal of Pharmacy Education*, *70* (5), 104.

Schafer, W. D., Swanson, G., Bene, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, *14*, 151-170.

Schamber, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. *Journal of General Education*, *14*, 151-170.

Schirmer, B. R., Bailey, J., & Fitzgerald, S. M. (1999). Using a writing assessment rubric for writing development of children who are deaf. *Exceptional Children*, *65*, 383-397.

Shadish, W., Cook, T., & Campbell, D. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessment in science. *Applied Measurement in Education, 4*, 347-362.

Stemmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, *36* (2), 102-107.

Suda, K., Franks, A., McKibbin, T., Wang, J., & Smith, E. (2008, July). *Differences in student performance based on student and lecture location when using distance learning technology.* Paper presented at the annual meeting of the American Association of Colleges of Pharmacy, Chicago.

Thaler, N., Kazemi, E., & Huscher, C. (2009). Developing a rubric to assess student learning outcomes using a class assignment. *Teaching of Psychology*, *36* (2), 113-116.

Toth, E. E., Suthers, D. D., & Lesgold, A.M. (2002). Mapping to know: The effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, *86*, 264-286.

Weiss, G. L., Cosbey, J. R., Habel, S. K., Hanson, C. M., & Larsen, C. (2002). Improving the assessment of student learning: Advancing a research agenda in sociology. *Teaching Sociology*, *30*, 63-79.

Wellman, J. V. (2001). Assessing state accountability systems. *Change*, *33*, 46-52.

Wiggins, G. (1998). *Educative assessment*. San Francisco: Jossey-Bass.

Acknowledgment

This research was supported by an ACL Grant from the Division of Academic Affairs, University of Alabama. Appreciation is extended to Dr. David R. Forde, for his insight during the model development phase of the research. An overview of this research was presented at the 2009 annual meeting of the Academy of Criminal Justice Sciences, San Diego, California.

Rebecca J. Howell *is an assistant professor of criminal justice at the University of Alabama. Her expertise includes learning outcome assessment, program evaluation, and the epidemiology and etiology of juvenile delinquency. Currently, she is conducting a funded outcome evaluation of the impact of service learning on students' sense of civic responsibility.*