

7-2009

A Method for Collaboratively Developing and Validating a Rubric

Sandra Allen

Columbia College Chicago, sallen@colum.edu

John Knight

University of Tennessee at Martin, jknight@utm.edu

Recommended Citation

Allen, Sandra and Knight, John (2009) "A Method for Collaboratively Developing and Validating a Rubric," *International Journal for the Scholarship of Teaching and Learning*: Vol. 3: No. 2, Article 10.

Available at: <https://doi.org/10.20429/ijstl.2009.030210>

A Method for Collaboratively Developing and Validating a Rubric

Abstract

Assessing student learning outcomes relative to a valid and reliable standard that is academically-sound and employer-relevant presents a challenge to the scholarship of teaching and learning. In this paper, readers are guided through a method for collaboratively developing and validating a rubric that integrates baseline data collected from academics and professionals. The method addresses two additional goals: (1) to formulate and test a rubric as a teaching and learning protocol for a multi-section course taught by various instructors; and (2) to assure that students' learning outcomes are consistently assessed against the rubric regardless of teacher or section. Steps in the process include formulating the rubric, collecting data, and sequentially analyzing the techniques used to validate the rubric and to insure precision in grading papers in multiple sections of a course.

Keywords

Rubric, Professional collaboration, Student learning assessment, Writing Public relations, News release writing

A Method for Collaboratively Developing and Validating a Rubric

Sandra Allen Columbia
College Chicago Chicago,
Illinois, USA
sallen@colum.edu

John Knight
University of Tennessee at Martin
Martin, Tennessee, USA
jknight@utm.edu

Abstract

Assessing student learning outcomes relative to a valid and reliable standard that is academically-sound and employer-relevant presents a challenge to the scholarship of teaching and learning. In this paper, readers are guided through a method for collaboratively developing and validating a rubric that integrates baseline data collected from academics and professionals. The method addresses two additional goals: (1) to formulate and test a rubric as a teaching and learning protocol for a multi-section course taught by various instructors; and (2) to assure that students' learning outcomes are consistently assessed against the rubric regardless of teacher or section. Steps in the process include formulating the rubric, collecting data, and sequentially analyzing the techniques used to validate the rubric and to insure precision in grading papers in multiple sections of a course.

Keywords: Rubric, professional collaboration, student learning assessment, writing, public relations, news release writing

Introduction

Preparing students for their careers is one goal of teaching and learning. Arming students with the skills and knowledge required for successful employment depends on congruence between professional standards and academic principles. At the same time, students deserve feedback on and assessment of their work that is valid and reliable regardless of the instructor and section of the course.

Rubrics are one way to structure the pedagogy for desired learning outcomes. Rubrics also provide an effective, efficient, equitable assessment method that can be understood and applied by both student learner and academic assessor. According to Stevens and Levi (2005), rubrics (1) provide a task description; (2) develop a scale for evaluation; (3) develop dimensions of the task necessary for success; and (4) provide a description of the dimensions relative to the scale being used.

This paper outlines an eight-step process for academics and professionals to collaboratively develop and validate a rubric that ties students' success in the academic world to their career readiness. We focused on (1) insuring the academic concepts and dimensions within the rubric are the learning skills students need; and (2) insuring those dimensions are professionally anchored.

We present eight steps as a guide that readers might adapt to meet their own needs. Within each step, we provide an example of how faculty at Columbia College Chicago collaborated with public relations professionals in Chicago to develop and validate a rubric for public relations writing.

Literature Survey

Rubrics can clarify an assignment and help students reach its learning objectives (Stevens & Levi, 2005; Schneider, 2006; Yoshina & Harada, 2007). In many cases, educators have attempted to develop rubrics as a systematic way to describe the tasks and specify what skills are needed. Some educators have developed rubrics to provide measurement guidelines to assess performance (Nagy, et al. 2000; AACSB, 2007) and facilitate giving timely feedback to students (Rucker & Thomson, 2003; Smith, 2008).

Andrade (2005) provides background on the structure and purpose of rubrics and catalogues their benefits as teaching and grading tools. To be effective, he cautions, rubrics must be valid, reliable and fair. For Messick (1989) validity is an evaluative judgment resting on the integration of evidence and theory to support inferences based on test scores. McNamara stressed the importance of *a priori* construct validity of the rubric where the pedagogy identified skills from professionals and other sources, and *a posteriori* construct validity relating to empirical and statistical validation.

McNamara (1996) suggests examining available literature to inform the content of the rubric. More significantly, he also suggests collecting information from experts to develop, determine, and test the rubric's specifications and scoring procedures. While Stevens and Levi (2005) believe collaborating to develop a rubric offers "an opportunity to discuss shared goals and teaching methodologies, and . . . evaluate and validate. . . grading practices" (p. 68), they also admit it is rare.

Literature abounds with information on methods to collect data from multiple sources. According to Helmer (1977), the Delphi method facilitates the formation of group judgment and has been used in education and in professional fields (Cornish, 1977). The Delphi method (Adler & Ziglio, 1996) is an exercise to collect and distill knowledge from a group. It is a means of asking participants a series of questions in two or more rounds. After each round, a facilitator collects individual's responses and their justification. The facilitator then reports a summary of the responses without identifying the sources. As the rounds continue, the responses are refined, eventually moving the participants to consensus.

Research is notably sparse on the collaborative process of developing and validating a rubric that integrates data collected from academics and professionals. Similarly there are but a handful of sources that address the importance of reliability in assessing students work regardless of teacher or section. This paper attempts to address these needs.

The authors also discuss the experience of collaboratively developing and validating an assessment rubric for a public relations writing course at Columbia College Chicago, a private Midwestern arts and media college.

Teaching writing skills is a critical element of student preparation for professional employment. The National Commission on Writing report titled, "Writing: A Ticket to Work ... or a Ticket Out" (College Board, 2004) called writing a threshold skill for employment. Public relations practitioners and educators agree: writing is fundamental to a student's academic success and professional career, and is the student's most marketable skill (McCleneghan, 2006; Cole & Corner, 2008; Steel, 2008). In the public relations field, Hardin and Pompper (2004) noted that practitioners lament expending significant resources to bring the writing skills of new hires up to par, particularly those skills needed to craft a news release.

Rios (2008) and Tucker, Derelian and Rouner (1997) recognize the news release as the time-honored convention combining critical thinking, writing and communication skills. Crafting the news release brings into practice basic news-writing skills which conform to conventional journalistic style. Writing the release involves several major activities, including generating and organizing ideas, translating those ideas into written form, and revising the written product (McMaster & Espin, 2007).

By using a collaborative method to develop and validate a rubric, the pedagogy and assessment of the students' work are grounded in the expectations of academia and the profession.

Method

Procedure

Chart 1 presents the development and validation process for a rubric that incorporates professional and academic input, and feedback and data analysis. It also illustrates the ongoing discussion, clarification, and incremental review of the pedagogy, including the specific collaborative approach for developing a rubric for news release writing at Columbia College Chicago.

The process is founded on the principle of continuous feedback and improvement. While the authors present the series of steps from beginning to end, we encourage readers to determine which steps may be useful in their situation. For example, those readers who have developed a rubric can extract and modify suggestions for analyzing its validity and reliability without necessarily retracing preceding steps. We hope our example provides useful information and guidance for the reader.

Table 1. 8 Steps in a Process to Collaboratively Develop and Validate a Rubric

	Description	Generalized model	Specific example of application at Columbia College Chicago
Step 1	Develop learning objectives for course	Discussion among faculty who teach course; input from professionals and the sources related to subject matter; research professional, certification requirements and exams	Discussions among faculty who teach public relations writing; input from professionals and academicians from other institutions; literature review, including information from professional organizations like Public Relations Society of America (PRSA)
Step 2	Identify sample of work to link learning	Review professional sources, including special reports, conference	Examined PRSA Port of Entry Report and PRSA requirements for professional accreditation; professional-source review;

	outcomes to professional competencies	proceedings, etc.; academic literature search	frequent meetings with and discussions among faculty teaching PR writing
Step 3	Develop evaluation rubric	Primary and secondary research; professional input; repeated discussions with faculty about work sample	Regular faculty meetings to discuss PRSA model and certification requirements, primary and secondary research; one-on-one meetings with academics; interviews with professionals
Step 4	Longitudinally test student learning as measured by rubric	Establish professional baseline; refine with academic and professional input	Two-test process: Test 1: PR professionals used rubric to evaluate student work sample; Test 2: professionals evaluated second sample of student work; faculty analyzed results
Step 5	Identify problems with sub-optimal performance	Use longitudinal data to assess rubric and pedagogy; evaluate uniformity in approach in multiple sections taught by many faculty	Faculty evaluated rubric against test results; then altered rubric's structure, expanded content, added detailed description to each of 5 categories; identified process to add weights to categories
Step 6	Improve construct validity of rubric	Determine weights for dimensions; collect and distill knowledge from professionals and academicians	Used the Delphi method first with professionals and then with faculty to reach consensus on weights in each category of the rubric
Step 7	Determine ability of rubric to differentiate between submissions	Assess rubric's ability to differentiate individual student work of varying performance; assess stability of rubric's standards; analyze variance to develop statistical evidence of rubric's accuracy and consistency	Faculty and professionals separately used rubric to assess and grade a weak and a strong news release; used two-way ANOVA to determine if significant differences exist; used F test for equal variances
Step 8	Analyze data to improve reliability and further validate rubric	Establish inter-rater reliability; determine a baseline; test for rater repeatability and reproducibility	Faculty graded one news release; used Delphi method to reach consensus on grade; used F test for equal variances; <u>looking ahead</u> ; faculty will grade same paper; we will then test for repeatability and reproducibility

Step 1: Develop Learning Objectives for the Course

Stevens and Levi (2005) call step one the "reflection" stage. In general, this step requires discussion and agreement on learning objectives among faculty. Step one also calls for input from practicing professionals whose expertise in the area provides real-time knowledge of the skill set students need to succeed in the profession.

Sources for this information are broadly available. Each academic discipline has related professional organizations that define requisite skills for certification. These certification requirements and exams provide a foundation for developing academic learning objectives, and reveal those areas with the greatest importance, or weight in the profession.

To develop learning objectives for the public relations writing course, faculty at Columbia College Chicago (some of whom are adjuncts who also are practicing public relations professionals) discussed course objectives among themselves. We called for input from academicians at other institutions, reviewed academic literature, and examined material from the Public Relations Society of America, the organization that accredits public relations professionals.

Step 2: Identify a Representative Sample of Work that Demonstrates Learning Competency

Not every course objective and concept can be covered by a rubric. In step 2 the goal is to identify an assignment that requires students to demonstrate learning competency in one of the main objectives of the course. This assignment must also require students to produce a work sample, which should incorporate the required skill set of the students' future career.

In this step, readers should review professional sources including special reports and conference proceedings for guidelines. An academic literature search also may yield content material. The faculty should discuss their findings, develop the assignment, and agree on and accept the work sample as the focus of their effort. As instructors collaborate to develop the assignment, their dialogue brings them to agreement on goals of the pedagogy.

The Public Relations Society of America's (PRSA, 2006) highly regarded Port of Entry Report provided significant direction for the 7 faculty who teach the public relations writing course at Columbia. Results of the literature search also informed our selection of a work sample. So, after many discussions, we elected to focus on the news release as the work sample.

Step 3: Develop an Evaluation Rubric for Uniformity of Teaching in All Sections

The authors found the work by Stevens and Levi (2005) useful as a basic primer in developing the rubric, and its dimensions and grading scale. In general, this step requires concerted leadership by a faculty member or program director who believes a uniform pedagogy is practical and achievable.

Essentially, the process calls for extensive input and discussion among the multiple instructors teaching the course and practicing professionals with their real-world perspective. In the process, several iterations of the rubric evolve as the differences of opinion are discussed and eventually resolved.

This stage of the rubric development began at Columbia in fall 2005 when a faculty committee met to: (a) develop a rubric based on the PRSA model; (b) review of existing literature; (c) conduct primary and secondary research; (d) meet one-on-one with academics; (e) interview with public relations professionals; and (f) meet regularly and informally with the PR Writing faculty.

In its first iteration (Appendix A), the rubric used a simple binomial, yes/no standard for evaluating five broadly defined categories describing acceptable news releases. In fact, the final analysis was merely an acceptable/not acceptable rating of the writing.

Step 4: Longitudinally Test Student Learning Against Professional Assessments

Step four is a two-test process. Test 1 establishes a baseline as a point of reference to measure and evaluate student learning over time. Over time may mean weeks or months, depending on the reader's individual situation. To establish a baseline, identify appropriate professionals to use the rubric to evaluate the students' work sample. Collect and retain the data.

After an appropriate interval of time, gather another sample of the students' work for test 2. Ask professionals to use the rubric to evaluate this sample. The reader now can comparatively analyze results of tests 1 and 2 to measure changes in student learning.

At Columbia, we established the initial baseline in summer 2006. One hundred twenty-two students were enrolled in 8 sections of PR Writing in the spring 2006 semester. We randomly selected 50 percent (N = 61) of their work samples for evaluation. Readers who are interested in the random selection process should know it wasn't perfect, but it worked. The names of the students were redacted from the news release. A student aide then sifted through the stack of news releases and eliminated 30 samples. From the resulting 92, he winnowed it to 61. A panel of 4 senior-level professionals from a prominent worldwide public relations agency used the rubric (Appendix A) to assess these 61 news releases.

In summer 2007, we repeated the process to longitudinal test for learning outcomes. Again, the work of 61 students was randomly selected by a different student aide. The same 4 professionals who did the initial benchmark in 2006 used the same rubric to evaluate these new samples of student work.

When the results of the baseline evaluation in 2006 were compared with the longitudinal test in 2007, we found a 10% increase year-over-year in the number of news releases rated acceptable work for entry-level public relations professionals. However, data also indicated that while 70% of the students evaluated in 2007 wrote acceptable news releases, the remaining 30% needed to further strengthen basic writing skills.

The faculty met to discuss the results of the evaluations. While the rubric seemed a useful tool for evaluating student work, we could not qualify acceptable and not acceptable work. In other words, the students seemed to be learning, but we could not tie the pedagogy to the results. We decided this iteration (Appendix A) of the rubric did not meet our needs.

Step 5: Identify Problems with Sub-Optimal Learning Performance

In step 5 readers should use the longitudinal data to assess the rubric's usefulness, and also to evaluate the instructor's approach to teaching the material. Consider how uniformly instructors use the rubric as a teaching, learning and assessment tool. For example, instructors may have differing interpretations of the terminology of the rubric's dimensions. Or instructors may emphasize different categories of the rubric for each assignment.

Based on the results of the longitudinal test, in fall 2007 the faculty at Columbia met to examine the usefulness of the rubric. We concluded the rubric's existing iteration, including the acceptable/not acceptable rating method, needed refinement to help us better evaluate the students' learning. We altered its structure, expanded its content, and added more detail to describe acceptable work in each of the 5 categories (Appendix B).

However, the faculty recognized the importance of first adding and then validating weights in each category. We wanted to be able to pinpoint where the pedagogy tied to student learning. We also wanted better evidence of how consistently their work met professionals' expectations of career-ready students.

Step 6: Improve the Construct Validity of the Rubric

After tying the basic validity of the rubric to academic and professional standards, improve validity by agreeing on the relative weights of the rubric's dimensions. Step 6 presents a method to determine these relative weights.

The Delphi method (Adler & Ziglio, 1996) is useful in collecting and distilling knowledge from a group. Essentially, it is a process where participants systematically in 2 or more rounds express an opinion on a problem or issue. For round 1, the facilitator poses the question, and solicits input from each participant in turn. After each round, the facilitator collects the feedback and reports out to the group without identifying which of the participants voiced a particular opinion. Based on this feedback, the participants are encouraged to revise their responses during the next round. As the rounds continue, the differences become smaller, until there is group consensus.

The authors endorse the Delphi method as an approach for developing consensus within groups. The experience at Columbia illustrates the application of the Delphi method. In 2007, 7 professionals ranging from assistant account executive (an entry-level position) to senior vice president at a highly regarded, worldwide public relations agency were solicited for their judgment. To insure broad-based input from public relations professionals, the individuals involved in this exercise worked at a different agency than those who evaluated learning outcomes in 2006 and 2007.

With the co-author as facilitator, the professionals met in a sixty-minute session. Each participant was provided with multiple copies of the rubric sufficient for several rounds. (Readers are directed to Appendix B, however, without the weights). Then we discussed the rubric and its 5 categories to ensure participants understood the content. The facilitator explained the objective of the exercise, that is, to determine the groups' best estimate of the relative value of each of the 5 categories of the rubric and to assign individual weights to each. The professionals were asked to assign points to each of the 5 categories in increments of 5 points that when added equaled 100 points total.

In round one, each participant independently and silently recorded the weights for each category. The facilitator collected the rubrics. The high and low values were presented to the professionals, and each high and low participant had the opportunity to justify his or her weighting. The facilitator led round 2. Participants again scored the 5 categories on a clean copy of the rubric. The facilitator collected the data and again reported out.

The exercise went 5 rounds with high/low weights given in Table 2.

Table 2. Delphi Method Results of Public Relations Professionals for Validating Weighting Factors

	Categories of the Rubric									
	Category 1		Category 2		Category 3		Category 4		Category 5	
Round	H	L	H	L	H	L	H	L	H	L
1	50	20	20	5	30	15	20	5	25	10
2	45	25	20	5	25	15	20	5	20	15
3	45	35	20	15	25	20	15	5	20	15
4	40	35	15	10	25	20	15	10	20	15
5	35	35	15	15	20	20	10	10	20	20

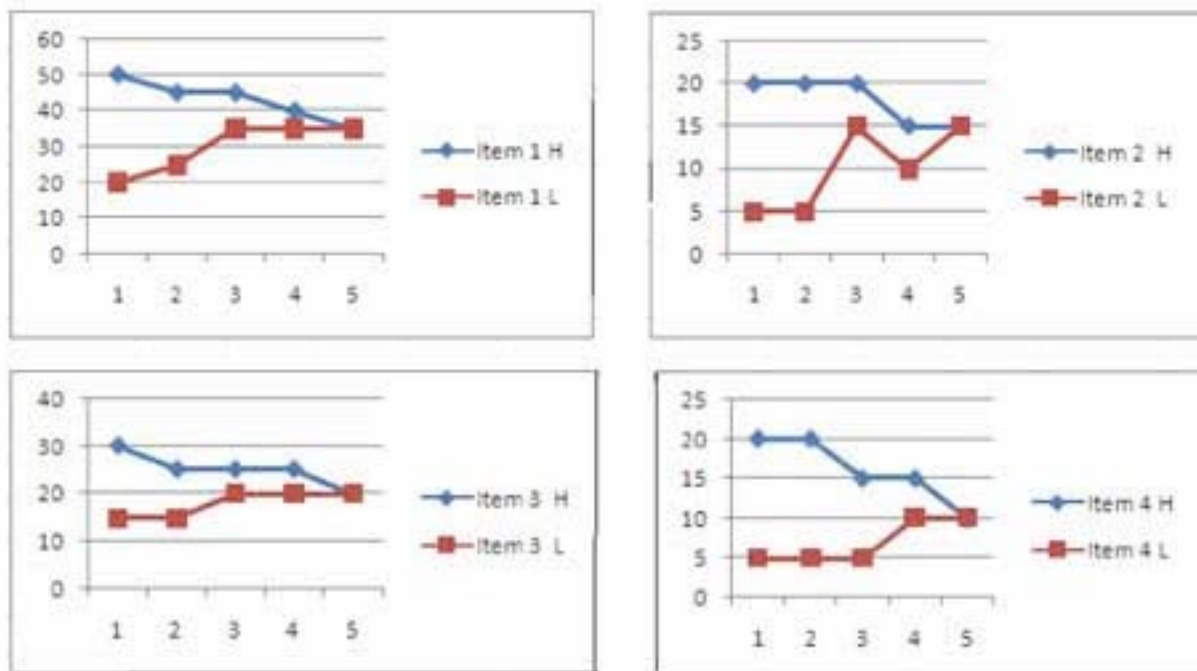
After reaching consensus, the professionals were asked to comment and give professional perspective as guidance for instructors who would use this rubric for teaching and evaluating students' news releases.

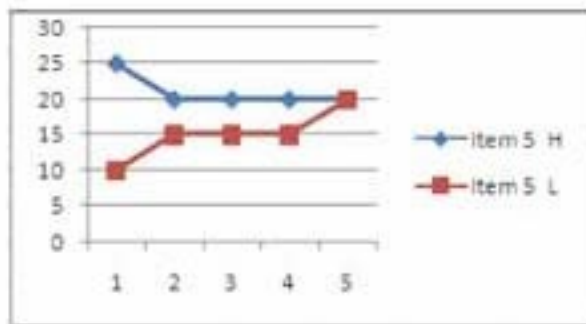
The participants stressed category 1, which requires information in the news release to be accurately written and effectively organized (#1 – Are the key ideas organized effectively?). "The message shapes the release," they agreed. "Without supporting points, the credibility (of the author and source of the release) is on the line." Thus, category 1 was most highly weighted with 35 points.

Categories 3 and 5 were equally weighted, with 20 points each. When discussing the weight of category 3, (In the headline/lead combination, is the news angle compelling?) the professionals agreed. "There's a short window to grab attention. The release must immediately establish that it has a compelling message. Otherwise, the media toss it out." The group also assigned 20 points to category 5 (Are the basics in place – grammar, sentence and paragraph mechanics?): "Spelling and grammar are wash-out errors. If these are wrong, then it's a no-go release."

The professionals considered categories 2 (Headline/lead combination) and 4 (Does the news release use a convincing journalistic style?) related to style, which in their professional opinions, were less important than substance.

Charts 1–5. Illustrations of the 5 Rounds of the Public Relations Professionals' Progression to Group Consen





Like the exercise with the professionals, faculty also used the Delphi method to reach consensus on the rubric's values and weights in its 5 categories. However, the exercise was carried out by email because the participants include adjunct faculty who were not available on the same day at the same time. Faculty were sent the rubric (Appendix B) without weights and given the same directions as the professionals. After each round, the facilitator collected the data and sent an email with the data to the faculty, sharing their remarks anonymously. At the conclusion of each round, the facilitator sent each of the participants a fresh rubric.

Though considerably more cumbersome than a face-to-face meeting, the back-and-forth messages built an email thread that provided a rich source of data for later analysis. The exercise went 5 rounds. Table 3 reflects the faculty's high/low weights for each round.

Table 3. Results for Faculty Using Delphi Method to Validate Weighting Factors

Round	Categories of Rubric									
	Category 1		Category 2		Category 3		Category 4		Category 5	
	H	L	H	L	H	L	H	L	H	L
1	25	15	30	15	25	10	30	20	30	10
2	20	15	30	15	25	15	25	15	20	15
3	20	15	30	20	25	20	25	20	20	15
4	20	15	30	20	25	20	25	15	20	15
5	15	15	25	25	20	20	25	25	15	15

After reaching consensus, faculty were asked to share their overall opinions. Faculty weights differed markedly from the professionals' results. Faculty considered category 2 (In the headline/lead combination, Is the message clear and compelling?) and category 4 (Does the news release use a convincing journalistic style?) the most important. Categories 1 and 5 received the lowest weighting. Table 4 outlines their differences in opinion.

Table 4. Summary Comparison of Weights Assigned Categories of a Rubric as Determined by Public Relations Professionals and By Faculty

Rubric Category	Professional	Faculty
Are the key items organized effectively?	35	15
In the headline/lead combination, is the message clear and compelling?	15	25
In the headline/lead combination, is the news angle compelling?	20	20
Does the news release use a convincing journalistic style?	10	25
Are the basics in place – grammar, sentence and paragraph mechanics?	20	15

The differences in and the justification for the final weights provided the basis for several conclusions. First, the professionals used a more “acceptable/not acceptable” logic to assign weights. Based on strengths and/or deficiencies in categories 3 and 5, the professionals considered the news release a “go/no go”. Faculty weighting showed less dispersion between categories, perhaps recognizing the students’ need to develop multiple skills in all categories of news release writing.

Secondly, though final convergence of the two groups was not as close as had been hoped, the differences did shed light on the areas that need to be explored for continuous improvement of the pedagogy. Delphi sessions with the faculty and professionals may encourage dialogue and mutual understanding of the specific terms of the rubric, which is important to the students’ career readiness.

Step 7: Determine the Ability of the Rubric to Differentiate Submissions to the Rubric Step 7 determines how well the rubric differentiates between students’ work, and also determines if academics and professionals agree on grading levels. The assessment of any one paper on the average should be the same when graded by professionals and by academicians, but demonstrate differences between individual student papers. Additionally, by analyzing the variance of grades assigned to any one paper, the precision of the assessment can be gauged by paper or by the group that assesses the paper.

At Columbia, we wanted to develop statistical evidence that the rubric can be accurately and consistently administered. In spring 2008 faculty who teach public relations writing submitted to the author 5 news releases from each of 8 sections (N = 40). Students’ names were redacted. The author reviewed all 40 news releases, and randomly selected 10 weak and 10 strong news releases and put them into 2 separate stacks. A student aide randomly selected one news release from each stack. The weak release was marked news release X; the strong marked news release Y.

Six of Columbia’s PR Writing faculty and 5 public relations professionals in Chicago were asked to use the rubric to assess news releases X and Y. Although either version of the rubric with its weights could have been used to insure comparability of final grades, we believed the faculty’s version (Appendix B) was more oriented to teaching and student learning. The faculty and professionals used this rubric to assign points in each category based on their assessment of the news release’s conformance with the rubric’s dimensions. Points were totaled.

We collected the results and used a two-way ANOVA to determine if, when we used the rubric, significant differences between news releases X and Y could be detected by the grader. We also wanted to determine if there were significant differences between how faculty and professionals awarded points to news release X and Y. Table 4 presents the average results of the final assessment.

Table 5. Average Results of Assessing and Awarding Points to 2 Different Student News Releases

	Points awarded news release X	Points awarded news release Y	Averages
Professionals (n=5)	45.4	77.3	61.35 N = 5
Faculty (n=6)	42.6	78.74	60.67 N = 6
Averages	43.9	78.05	60.97

We then did a two-way ANOVA (Table 6).

Table 6. Results of Two-way ANOVA

Two-way ANOVA: Grade versus Fac. vs. Prof., X or Y					
Source	DF	SS	MS	F	P
Fac. vs. Prof.	1	2.8	2.80	0.01	0.910
X or Y	1	6915.6	6915.61	32.69	0.000
Interaction	1	25.2	25.22	0.12	0.734
Error	20	4230.7	211.53		
Total	23	11174.3			

S = 14.54 R-Sq = 62.14% R-Sq(adj) = 56.46%

Statistically significant differences ($p = .000$) resulted when news releases X and Y were assessed. However, no significant differences existed between the average actual grading practices of the professionals and the faculty ($p = .910$). Also, no interaction was found within the model ($p = .734$). Thus, we had confidence that using the rubric to assess different news releases differentiates between the quality of the students' writing. The results also indicate faculty and professionals are similar in their assessment of the quality of different papers.

Further, the ANOVA results indicate that an individual student submitting a news release to an individual grader who uses the rubric would be subject to a precision of ± 14.54 points with 68% confidence, and approximately ± 29.32 with 95% confidence given the overall standard deviation from the ANOVA of 14.54 (see $s = 14.54$ in Table 6). The precision in the estimates indicate that additional analysis of the assessment process is warranted.

Because faculty ultimately will grade the students' news releases, to limit unnecessary variance contributions from the professionals and analysis of the variances of grades from only the faculty was performed. An F test for equal variances for each news release using only the faculty grades indicated that significant differences existed in the variances of the grades assigned to each paper ($p = .044$). The sample variance for news release X was 8.42 ($s = 14.36$) while the sample variance for news release Y was 3.05 ($s = 5.20$).

The data on the best estimates of the standard deviations of both news releases as graded solely by faculty indicate that the higher quality news release Y has a smaller standard deviation. News release Y with an average of 78 points has a best estimate standard deviation of only 5.2 which indicates that a strong news release could be reasonable graded within ± 5.2 points (one letter grade) in 68% of the cases. However, news release X, the weaker news release, has a best estimate standard deviation of 14.36. Thus, a relatively weak news release would have a grading range of about 14.36 with 68% confidence (possibly 3 letter grades) depending on the faculty member.

The analysis led to several observations. We believe, without numerical support, as a news release improves and approaches a higher average, given the 100-point scale the variance will necessarily contract. It's possible that higher quality is simply easier to discern and evaluate. Faculty who evaluate weak news releases also may be reluctant to discourage (rather than encourage) student learning. From a student perspective, their grade would be highly teacher dependent. Additionally, the analysis indicated the possibility that the scale for grading each element of the rubric is subject to too much interpretation, especially when poor papers are evaluated. For continuous improvement, we intend to add a more defined numerical and descriptive scale to each dimension of the rubric.

As an additional result of our analysis, faculty will require students to rewrite weak news releases, and submit for re-evaluation. We believe rewriting to improve quality may facilitate student learning and provide a more precise assessment of their work. Not incidentally, as the news release improves, the variances in results measured by the rubric may diminish.

Step 8: Improve the Reliability of the Rubric with Data Analysis to Further Validate the Rubric

In step 8, the faculty who teach multiple sections of the course must commit to improving the reliability of the rubric, its sub-elements and terms, and in applying its grading scale to improve inter-rater reliability.

This paper is not intended to address all possible quantitative tools. However, readers might first establish a baseline for analysis. Once a quantitative baseline is established, opportunities for improved reliability are significant if subsequently tested against the baseline. To increase overall reliability of the rubric, both repeatability and reproducibility must be measured. After repeatability has been determined and measured, reproducibility then can be more explicitly measured. Reproducibility is the degree to which differences occur between grades assigned by different instructors on the same paper.

At Columbia, the faculty met in December 2008 to improve the reliability of grading across the 13 sections of PR Writing taught by 9 faculty. The faculty used the rubric to evaluate 1 news release written by 1 student. We used the Delphi method to discuss and defend high and low scores. Faculty agreed on and defined standards for awarding points in each of 5 categories of the rubric. In round 2, the 9 faculty again evaluated the same paper to determine if we could reach greater uniformity. In round 3, we reached consensus (Table 7).

Table 7. Results of Delphi Method to Demonstrate Reliability in Assessing Student News Release

Instr.	1	2	3	4	5	6	7	8	9	Ave.	S.Dev.
Round											
1	79	100	79	85	88	84	82	85	91	85.89	6.57
2	83	94	84	85	88	85	86	87	90	86.89	3.41
3	88	88	88	88	88	88	88	88	88	88.0	

An F test for the differences in variances between rounds 1 and 2 had a p value of .082, indicating a significant reduction in the variance of grading at the 10% level. The precision of the estimate of differences in reliability of grading is approximately ± 6.8 with 95% confidence. Given the small sample size ($N = 9$), the results indicate that reliability improves when faculty examine the data, and then discuss and resolve differences in interpreting terms and applying standards. Readers should note, however, in actual practice and across many papers whose quality varies from strong to weak, the precision likely will not be as marked.

Looking ahead, the authors intend to build upon this foundation to test for rater repeatability and reproducibility, with the ultimate goal of increasing the precision of grading among the different instructors. Our research contemplates using the Delphi method to develop more explicit relational and ordinal grading scales for each sub-category of the rubric. We intend to longitudinally test for repeatability, that is, whether a particular grader can assign the same score to the same paper twice. We know this is critical to establishing gauging capabilities of the rubric and identifying the final percentage of variation in grading that can be assigned to inter-rater differences.

In late 2009 to assess repeatability, the faculty who teach public relations writing will independently evaluate the same set of 10 different news releases written by 10 different students. Before the end of the year, but when sufficient time has elapsed to forestall remembering each score, the faculty will re-evaluate these 10 news releases. To establish a new baseline of the precision of the rubric, an ANOVA will determine if significant differences exist in the average scores awarded by the different instructors. If significant differences exist, we will discuss, review and make clear the issues that differentiate the faculty's interpretation of the rubric. Our ultimate goal is to assure we can determine a reasonably precise and reliable score for any news release regardless of the instructor. We want the student to have confidence in the assessment process regardless of teacher or course section.

Discussion

The authors wanted to develop and outline a collaborative process for academics and professionals to develop a rubric that was both flexible and straightforward. We wanted a process any instructor or group of instructors could use regardless of the course material. The 8 steps in the rubric development and validation process are not prescriptive. The steps can be generalized to fit the simplest of needs, or embellished and extended with additional data to fit more specific situations. Each of the 8 steps can stand alone. Fundamentally, though, each step incorporates the principle of continuous improvement based on analysis and on feedback from faculty and professionals.

As Stevens and Levi (2005) state, the process is neither complicated nor unduly time consuming, and benefits of collaborating are available to all participants. The authors found the incentive for faculty is the promise of a sound pedagogy. Professionals benefit by having

available a pipeline of career-ready graduates. The advantage for the student is clear, fair feedback linked to the expectations of current practitioners' in the student's future profession. To be career-ready, to learn what's important for success, students deserve no less from their instructors.

Acknowledgements

The authors wish to recognize and thank the two groups who made this study possible: (1) The faculty who teach Public Relations Writing at Columbia College Chicago: Paul Berner, Dominic Calabrese, Matthew Carlson, Richard Greb, Michael Kooi, Joseph Marconi, Anne Marie Mitchell, Mark Perlman, and Cynthia Pike-Fuentes; and, (2) The public relations professionals at Hill & Knowlton Public Relations Chicago and Fleishman Hillard Public Relations in Chicago. We also acknowledge initial work on teaching news release writing and assessing student learning outcomes by colleagues Alton Miller and Laurence Minsky.

References

- Adler, M. & Ziglio, E. (1996). *Gazing into the oracle*. Bristol, PA: Jessica Kingsley Publishers.
- Andrade, L. (2005). Teaching with rubrics: the good, the bad, and the ugly. *College Teaching*, 53, 27-30.
- AACSB. (2007). *Assurance of learning standards*. Report by the Association to Advance Collegiate Schools of Business
http://www.aacsb.edu/accreditation/process/documents/AACSB_STANDARDS_Revised_Jan08.pdf
- Bloom, B.S. (1956) *Taxonomy of Educational Objectives, the classification of education goals – Handbook: Cognitive Domain*. New York: McKay.
- Cole, R. & Corner, E. (2008) How are we doing? Time for research on teaching writing for public relations. *TACTICS*, 18-20.
- College Board (2004) *Writing: A ticket to work ... or a ticket out: A survey of business leaders*. Retrieved January 27, 2007, from
www.writingcommission.org/prod_downloads/writingcom/writing-ticket-to-work.pdf
- Cornish, E. (1977). *The study of the future*. Washington, D.C.: World Future Society.
- Hardin, M. & Pompper, D. (2004). Writing in the public relations curriculum; practitioner perceptions versus pedagogy. *Public Relations Review*, 30, 357-364.
- Helmer, O. (1977) Problems in futures research: Delphi and causal cross-impact analysis. *Futures*, 17-31.
- McCleneghan, S. (2006). PR executives rank 11 communication skills. *Public Relations Quarterly*, 51, 43-45.
- McMaster, K. & Espin, C. (2007) Technical features of curriculum-based measurement in writing. *The Journal of Special Education*, 41, 68-84.
- McNamara, F. (1996). *Measuring second language performance*. London: Longman.

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

Nagy, P. (2000). The three roles of assessment: gatekeeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(2), 262-279.

Rios, M. (2008). Following communication rules: A communication-centered theory for public relations. In T. Hansen-Horn & B. Neff (Eds.), *Public relations: From theory to practice* (pp.181 – 192) Upper Saddle River, NJ: Pearson Education, Inc.

Rucker, M.L., & Thomson, S. (2003). Assessing student learning outcomes: An investigation of the relationship among feedback measures. *College Student Journal*, 37 (3), 400- 405.

Schneider, F. (2006). Rubrics for teacher education in community college. *Community College Enterprise*, 12, 39-55.

Smith, L.J. (2008). Grading written projects: What approaches do students find most helpful. *Journal of Education for Business*,83(6), 325-330.

Steel, A. (2008, Feb. 4). Careers: Expert Advice. *PR Week*, 19.

Stevens, D.D. & Levi, A. (2005). *Introduction to Rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning*. Sterling, VA: Stylus Publishing.

Tucker, T., Derelian, D., & Rouner, D. (1997). *Public relations writing: An issue-driven behavioral approach*. Prentice Hall, Saddleback, New Jersey, 90-134.

Van Slyke, J. (Ed.). (2006). *Public Relations Society of America: The report on the commission on public relations education. Public relations education for the 21st century: the professional bond*. New York City: Public Relations Foundation.

Yoshina, J.M., & Harada, V. H. (2007). Involving students in learning through rubrics. *Library Media Coordinator*, 25(5), 10 -14.

Appendix A

Rubric for evaluating news release: Iteration #1

Student name: _____ Rating: Yes or No

<p>Are the key ideas organized effectively?</p> <ul style="list-style-type: none"> Points in the release support the main message <p>In the headline/lead combination, is the message clear and compelling?</p> <ul style="list-style-type: none"> Headline summarizes main point of release <p>In headline/lead combination, is the news angle compelling?</p> <ul style="list-style-type: none"> Lead 'graph includes 5Ws – who, what, when, where, why <p>Does the release use a convincing journalistic style?</p> <ul style="list-style-type: none"> Release conforms to Associated Press Stylebook standards <p>Are the basics in place?</p> <ul style="list-style-type: none"> Correct spelling, grammar, punctuation <p>Overall, is this news release acceptable?</p>		
---	--	--

Appendix B

Rubric for evaluating news release: Iteration # 3 (Faculty version)

News Release Evaluation	Points Earned/Avbl.
<p>1. <i>Are the key ideas organized effectively?</i> Do the points in the release support the main message? Does the quote in second or third paragraph move the story forward? Is the attribution correct? Do the paragraphs have one dominant idea? Overall, is the release well written?</p>	<p>___ of 15 points</p>
<p>2. <i>In the headline/lead combination:</i> Is the message clear and compelling? Is the headline a complete sentence? Does the tense in headline match tense of news release topic? Does the headline summarize main point of release? Does the headline "grab" attention? Does the headline summarize information in lead paragraph?</p>	<p>___ of 25 points</p>
<p>3. <i>In the lead:</i> Is the news angle compelling? Is information timely? Does the first sentence establish local interest? Does the first sentence introduce the news angle? Does the lead paragraph includes 5 W's</p>	<p>___ of 20 points</p>
<p>4. <i>Does the news release use a convincing journalistic style?</i> Does the release conform to AP Stylebook standards? News is written in inverted pyramid fashion Least important information is in last paragraph</p>	<p>___ of 25 points</p>

<p>Does the release conform to standard format</p> <p>Dateline: Includes date Capital Letters Location of story</p> <p>Contact information: Includes name, title Day and night telephone numbers</p> <p>Second/third pages include slug line Pages end in (more) or ### Is the boilerplate appropriate? Includes source for more information and contact info</p>	
<p>5. <i>Are the basics (grammar, sentence and paragraph mechanics) correct?</i></p> <p>Are words spelled correctly? Is grammar correct? Is punctuation correct? Are sentences complete? Are sentences short and declarative? Is the release factually accurate?</p>	<p>___ of 15 points</p> <hr/> <p>TOTAL POINTS:</p>

©Sandra Allen 2009