Developments and Changes Resulting from Writing and Thinking Assessment

Teresa Flateby
University of South Florida

Abstract

This article chronicles the evolution of a large research extensive institution's General Education writing assessment efforts from an initial summative focus to a formative, improvement focus. The methods of assessment, which changed as the assessment purpose evolved, are described. As more data were collected, the measurement tool was transformed into a system of assessment. Additionally, challenges encountered are discussed.

Introduction

Ten years ago the General Education assessment team at the University of South Florida (USF) used a holistic scale to evaluate student writing in the General Education curriculum. Student writing samples were collected at three points in the curriculum: when students (a) entered as first-year students, (b) completed their first year and, finally, (c) completed all general education courses. Raters who also scored the State's "rising junior" essay tests assigned scores of one through six connoting proficiency levels from "below" to "exceeds" expectations. While the results confirmed anecdotal evidence that some students were more than acceptable writers, they also indicated that many students were not proficient. Although we used this approach for several years (and collected summative data), we lacked formative data to identify specific student writing strengths and weaknesses that could inform instruction or the curriculum. Data confirmed writing deficiencies, but were not valuable for making changes and improvements, one essential purpose of assessment. As a result, the assessment team suggested evaluating the usefulness of an analytic rubric designed at USF for the classroom to address program assessment purposes.

The classroom rubric, developed before the formal assessment of General Education occurred, was initiated to address needs identified in a two-year, team-taught writing-intensive learning community program at the University of South Florida. One of the goals of this program was to encourage the deeper learning often associated with writing. Two discoveries led to the development of the rubric. The program coordinator, who is also a faculty member of the English department, and I (the external evaluator of the program) determined through interviews and surveys that grading of students' writing assignments varied widely among faculty. Also early and throughout the two-year program, we observed complex thinking through classroom observations, reflecting the upper levels of Bloom's Taxonomy of Educational Objectives—Cognitive Domain (1956). Responding to these two findings, we recognized the need for a tool that enables the consistent evaluation of students' writing and thinking skills by faculty from diverse disciplines. We reviewed existing performance-based measures, but did not ind any that fulfilled the identified needs. Thus, we began the development of the Cognitive Level and Quality of Writing Assessment (CLAQWA) rubric.

Based upon commonly used writing handbooks, such as *St. Martin's Handbook, Harbrace College Handbook*, and *Scott Foresman Handbook for Writers,* the initial writing rubric included a five point scale with only levels one, three, and five defined. The sixteen trait analytic rubric was organized into categories, which were modified after meeting with teams of faculty and applying the rubric to papers. Due to a writing style often observed in beginning students' essays, the single category "Organization and Development" was divided into two: one pertaining to structure and another reflecting reasoning and evidence supplied. We realized that while many beginning students' essays had an appealing structure (five paragraph essays that students learn to produce for standardized testing), the quality of content and quality of reasoning exhibited were often weak. These and other results were used to refine the rubric to represent the full range of writing – qualities associated not just with learning to write, but also writing to learn.

When searching for a framework for the thinking portion of the resulting two-part scale, we chose Bloom's Taxonomy of Educational Objectives-Cognitive Domain (1956). In addition to its accessibility, the taxonomy reflects the type of thinking faculty typically advocate, such as analysis, synthesis, and evaluation. Moreover, several authors have recommended this taxonomy to assess writing. In 1983 Spear

advocated the use of Bloom and his colleagues' work for writing evaluation, and Olson (1992) developed a writing curriculum around Bloom's cognitive levels. In 1997 Steele, in his rationale for the development of American College Testing's Critical Thinking Assessment Battery, (which required writing) maintained that "Bloom's Taxonomy remains useful as a means of analyzing and classifying the levels of intellectual demands in cognitive activities" (p. 19).

The work of Madaus and his colleagues (1973) provided the basis for USF's cognitive scale. Their work showed a branching at the higher end of the taxonomy, thus transforming it into a four-level taxonomy (instead of the original six-levels). We subdivided these four taxonomy levels into low, medium, and high categories. Unlike the writing scale, we chose not to define the categories within levels, because when using the cognitive scale to assess levels reached in student texts, we found little variation in instructors' judgments.

When first applying the rubric for program assessment purposes, we used the initial iteration of the scale (five levels, with levels one, three, and five defined). It soon became evident, however, that all five levels needed clear definitions to achieve acceptable inter-rater reliability. Indeed, if raters within the institution cannot agree on ratings of essays then it is impossible to make defensible statements about students' performance levels or to make comparisons over time, across years, or within groups. Thus, we began the laborious task of clearly describing all five levels of the sixteen element analytic scale.

This continuing phase of development underscores the evolutionary nature of rubric development and use. As data were gathered, variations and perceptions of the definitions surfaced. Because rubrics are based upon language, users' experience and biases, these factors impacted the interpretation of levels of the traits. As calculated by the percent of adjacent-rater agreement, acceptable inter-rater reliability values (.89-.93) were achieved following clarification of the rubric (Micceri, unpublished institutional document, http://usf.edu/assessment).

As we proceeded with the assessment of writing and thinking, we continued to collect data at the same points in the curriculum: the beginning of Composition 1, the completion of Composition 2, and in liberal arts "exit" classes that are completed in the junior and senior years. With this data collection plan we were attempting to ascertain if students were reaching expected writing levels and on which of the components of the writing rubric needed the most improvement. In collecting data, we randomly selected sections from Composition 1 and 2 classes and used essays from all students in those sections.. The data collection for exit classes was less structured; faculty volunteered to provide their sections' essays. Because the interest was in students' performance after completing the General Education curriculum, and not growth in these exit classes, this type of sample selection seemed defensible. We attempted, however, to ensure that students in the sample were representative of the relevant demographics of the USF student population.

Using Results

After scoring our students' essays with the analytic rubric for approximately three years, we made valuable discoveries, which were used to suggest instructional and curricular changes. For example, when we began measuring the cognitive levels reached in our junior and senior undergraduate students' texts, we developed a standard prompt within courses and allowed students a week to complete the assignment. Although written to elicit Level Four on the Cognitive Scale, results showed that student performance was lower than desired. This finding was consistent with the "Reasoning" and "Quality of Evidence" performance levels of the writing rubric. We were uncertain however, if students' performance was truly reflective of their achievement levels or if it was adversely affected by the prompt, which was only minimally tied to course content.

Due to this concern, we changed our assessment strategy to include assignments on instructors' syllabi, if they targeted sufficiently high cognitive levels. With this approach, we hoped to determine if connecting the prompt more specifically to class assignments would elicit higher thinking skills. Although not systematically researched, we made a significant discovery: the importance of the prompt. Faculty routinely thought they were asking students to write at higher cognitive levels than their prompt reflected, and often the expectations were unclear to students. In addition, after evaluating hundreds of students' papers written to address many different prompts, scoring teams found the prompts to be critical, not only for eliciting a specific cognitive level, but also clarifying expectations for students. More open-ended or ambiguous assignments produced lower performance than assignments with clear expectations. This finding has had broad-based instructional and faculty development relevance.

Our data and process revealed that even if faculty and assessment teams do not evaluate students' cognitive levels reflected in their writing, the conscious selection of appropriate cognitive levels and careful construction of the assignments to reflect these levels are important to eliciting desired writing. Also, attention to the cognitive levels helps ensure compatible results if comparisons are to be made. Our data support composition literature suggesting that when students begin writing at higher cognitive levels, often their writing skills deteriorate (Schwalm, 1985). This finding has both pedagogical and assessment implications. If a goal is for students to clearly communicate higher order thinking, they must be given adequate opportunities in multiple classes to develop these more advanced thinking skills. Also, for assessment purposes, an institution or program must decide which cognitive levels should be addressed in assignments, especially if comparisons are made; this too underscores the importance of carefully planning the assignment's cognitive level.

Another finding was used to make curricular changes. Results confirmed that many of our students were not writing at the level expected; more importantly, we discovered that the weakest areas pertained to thinking, such as providing supporting evidence, and developing and organizing ideas. Writing skills such as grammar and mechanics, while below desired levels, were stronger than critical thinking skills.

After assessing general education learning outcomes for several years, general education became the focus of our Quality Enhancement Plan, a plan required by the Southern Association for Colleges and Schools for improving student learning outcomes. The assessment data helped guide revisions to the general education curriculum, resulting in specific changes to address weaknesses discovered in students' writing and thinking. Process writing (encouraging revisions facilitated by feedback) is now required in four of the twelve general education courses. Central to the writing emphasis is the development of ideas, inclusion of supporting evidence, logical progression of ideas and cohesiveness of texts. In addition, the plan promoted graduate and undergraduate student training to assist with writing assessment and to provide feedback to larger classes. Another change introduced is a capstone course in which writing in students' disciplines is emphasized. Equally important, the general education curriculum now emphasizes critical and higher order thinking, as well as inquiry-based learning approaches.

In addition to the direct evidence collected, we gathered indirect survey data. These results indicated that some faculty were concerned about students' writing performance levels, felt ill-equipped to provide adequate feedback, were concerned about class sizes prohibiting the ability to give feedback, and were unsure if sufficient resources were available to help students with writing deficiencies.

To address some of these concerns, we have transformed our classroom and program assessment rubric into an online system (CLAQWA Online). This online system assists faculty, students, and assessment professionals to evaluate student writing and thinking across the curriculum and helps close the assessment loop. Faculty or assessment teams are able to select writing and thinking components appropriate for a particular assignment. The instructor or the team evaluates students' writing and thinking by indicating directly on students' online texts which of the five levels described for each element reflects the text and by providing additional comments, if desired. Students are then able to access their work, which have the weak or strong writing element levels embedded in their texts. Students are able to review online instructional examples written for all levels of each trait, with feedback explaining why each example represents a specific level. This review helps them understand performance at each level and improve their writing on a trait (thus closing the loop). Designed to aggregate results, faculty and assessment teams are easily able to determine problem areas to address in their classes or in programs, again helping to improve students' writing and thinking (i.e. close the assessment loop). Through the online system students are able to give feedback to each other, thus further engaging them in the writing and improvement process (http://www.usf.edu/assessment/CLAQWA/Online).

Also through our assessment processes we discovered another method for improving student writing, which has become valued by faculty. Several members of the scoring team who were teaching composition decided to modify the paper version of the CLAQWA rubric for peer review use in the classroom. Although peer review was already part of their classes, they found that the modified rubric produced improved writing as compared to the peer review process they had been using. The success experienced with peer review in composition classes led to questioning its applicability in classes from different disciplines. We have conducted several peer review studies to determine if improvement could be measured. In electrical engineering and literature classes, improvement was observed with paper or online approaches. In the most recent studies, focusing on peer review through the online system, measurable improvements

were found in varying degrees in all sixteen of the rubric's elements.

Challenges and Conclusions

Several challenges associated with the writing assessment are currently being addressed at the University. Although we made changes in the General Education curriculum in response to the assessment data, the actual instructional changes are not as widespread. Because the use of assessment results and faculty development opportunities are interdependent, identifying the person(s) or unit(s) responsible for coordinating results with development is critical. Without this coordination, the optimal use of assessment data may not be realized, which is often cited as an assessment weakness. The question of who is responsible for ensuring that data are actually used, especially for a general education curriculum, must be clearly established and faculty development opportunities must be directly tied to assessment results.

Related is the importance of administrators' support of these assessment efforts and the insurance that resources and rewards are available to faculty for making instructional and curricular changes based on assessment data. Gaining an administrative commitment may be difficult in some institutions, but is essential for promoting the message that assessment not only is essential for accreditation, but also for improving (maximizing) student learning.

Another finding relevant to other institutions' assessment processes is the importance of developing detail in rubrics. A rubric should provide clear operational descriptions associated with different levels of proficiency. For example, the criteria for paragraph construction that exceeds expectations is much clearer to faculty and students when a rubric uses language such as, "Each paragraph is unified around a topic that relates to the main idea. All paragraphs support the main idea and are ordered logically" rather than with simply "Exceeds Expectations." Furthermore, faculty tend to rate more consistently with each other when definitions are clearly articulated. Finally, we discovered that after these rubrics were fully developed that we were able to engage students in their own learning, improve students' writing and thinking, and demonstrate this improvement.

In sum, USF has learned a tremendous amount about its students' writing and has used this information to improve the quality of our instruction. To get to this point, however, required several years of careful thinking about how USF wants students to write, how to elicit this type of writing, and how to accurately assess it. That said, improving writing and its assessment at USF is still evolving. Implementation of the program could be more pervasive and support more robust. Persistence and sending clear messages to faculty and educating administrators that improving student learning is assessments' fundamental purpose may help diminish these challenges.

References

American Philosophical Association. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, ("The Delphi Report"). (ERIC Document Reproduction No. ED 315 423).

Bloom, B. S. (ed.) (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook 1: Cognitive Domain. New York: McKay.

Madaus, G. F., Woods, N. E. & Nuttal, R. L. (1973). A causal model analysis of Blooms' Taxonomy. *American Educational Research Journal, 10*, 253-262.

Micceri, Ted (2006). Unpublished institutional document, http://usf.edu/assessment.

Olson, C. B. (1992). *Thinking/writing: Fostering critical thinking through writing*. New York, NY: Harper Collins Publishers.

Schwalm, D. E. (1985). Degree of difficulty in basic writing courses: Insights from the oral proficiency interview testing program. *College English, 47*(6), 629-640.

Spear, K. (1983). Building cognitive skills in basic writers. *Teaching English in the Two-Year College*, *9*(2), 91-98.

Steele, J.M. (1997). *Identifying the Essential Skills in Critical Thinking at the Post-secondary Level to Guide Instruction and Assessment.* [Draft]. Iowa City: American College Testing.

CLAQWA information and website may be found at http://usf.edu/assessment