

## *Assessing Undergraduate Writing In Psychology*

*James F. Sanford, Ph.D.  
Terry Myers Zawacki, D.A.*

Office of  
Institutional Assessment  
D111 Mason Hall  
MS 3D2  
703-993-8834  
[assessment@gmu.edu](mailto:assessment@gmu.edu)  
<http://assessment.gmu.edu>

Director  
Karen M. Gentemann, Ph.D.  
[genteman@gmu.edu](mailto:genteman@gmu.edu)

Assistant Director  
Ying Zhou, Ph.D.  
[yzhou@gmu.edu](mailto:yzhou@gmu.edu)

Programmer/Analyst  
Ramesh Gali, M.S.  
[rgali@gmu.edu](mailto:rgali@gmu.edu)

Program Technician  
Anne Rose, B.S.  
[arose2@gmu.edu](mailto:arose2@gmu.edu)

---

*Writing assessment at George Mason is tailored to meet the criteria and expectations of faculty in each discipline. This publication was adapted from the edited summary of a paper presented by James Sanford, Associate Professor of Psychology at the "National Institute on the Teaching of Psychology" in St. Petersburg Beach, Florida, January 3-6, 2004. The paper was written by James F. Sanford, Ph.D. and co-authored by Terry Zawacki, D.A., Director of Writing across the Curriculum and the University Writing Center. Questions regarding this subject may be directed to the authors and to Karen Gentemann, Director of Institutional Assessment.*

Experts in writing assessment have long suggested that assessment of student writing be qualitative, not quantitative, and that it be scored impressionistically (Cooper, 1977). Impressionistic scoring means that papers are read quickly and evaluated according to general content, tone, argument, etc. Frequency counts and other micro-assessment techniques are avoided. According to Cooper, inter-rater reliability may be as high as +.90 when evaluators use holistic writing assessment procedures.

Clearly, evaluators must be well-trained if holistic assessment is to be valid. One specific holistic evaluation procedure that Cooper (1977) calls *Essay Scale* involves raters first extensively reviewing several essays holistically and discussing general impressions among themselves. From this review and discussion among raters, a small number of criteria for assessment and their operational definitions evolve.

After the raters are satisfied that the criteria are appropriate, student essays are reviewed and rated on each criterion. Cooper recommends that different criteria be developed for each general type of writing that is assessed. Overall, if assessors are well trained and the criteria are relevant and operationalized so that all assessors are judging quality of student papers in the same way, a valid holistic assessment will be completed.

In 2000, the State Council of Higher Education for Virginia mandated that all state colleges and universities develop and implement writing assessment. Colleges and universities were allowed flexibility in developing their own procedures. At George Mason University, the Director of the University Writing Center, working with the Writing Assessment Group and other faculty experienced in writing assessment, developed a holistic procedure that allowed writing assessors in each program to develop and evaluate student writing in a way that was appropriate to the needs and expectations of that program. That allowed English students to be evaluated on their competence in writing literary analyses, biology students in their competence in creating lab notes and modifying them into scientific summaries, etc.

In psychology, a sample of final papers in students' required research methods course was selected for writing assessment. These papers summarized final proposals for hypothetical research studies. The research methods course is a writing intensive course that includes draft review and revision. The final papers included an abstract, a literature review, the development of research hypotheses, a complete method section, hypothetical results and conclusions, and references. A random sample of 32 papers was selected from those that students submitted in fall semester, 2001.

## *Method*

Two faculty, the Chair of the Department of Psychology's Undergraduate Program Committee and one member of that committee, attended an all-day workshop that was led by two experts in holistic writing assessment. The workshop included faculty from across the curriculum and involved holistic assessment training. The workshop revealed the varying disciplinary assumptions that faculty bring to their definitions of what constitutes good writing. After completion of the workshop, a similar workshop was held to instruct psychology faculty in the assessment methods to be used. The workshop in the Psychology Department was led by the Director of Writing across the Curriculum with input from the two psychology faculty who had attended the prior workshop. Eight faculty volunteers participated. The workshop resulted in development of criteria and a scoring sheet for assessing psychology student writing. Following completion of the workshop, the eight faculty were divided into four pairs. Each pair was randomly assigned to evaluate eight student papers. First, each faculty member assessed four papers and assigned a score to each criterion. Then, papers were exchanged within pair members. Finally, any discrepancies between pair members in the assigned ratings were discussed and resolved.

## *Results and Conclusions Concerning Student Writing*

The 32 papers rated over five criteria yielded a total of 160 ratings. Of these, 39 ratings were "more than satisfactory," 90 were "satisfactory," and 31 were "unacceptable." Thus, the large majority of ratings (80.6%) indicated that psychology students demonstrated at least acceptable performance, while less than one-fifth (19.4%) were unacceptable to faculty evaluators.

Results were examined in more detail in two different ways. First, a frequency distribution of the overall distribution of ratings was computed (see Table 1). From this table, it can be seen that two students earned "more than satisfactory" ratings in all five evaluation categories, while, at the negative end of the distribution, two students earned one "satisfactory" and four "unacceptable" ratings. Other observations from Table 1 include (a) 16 students earned at least one "more than satisfactory" rating, while 16 students also earned at least one "unacceptable" rating; and (b) individual student ratings were generally consistent across criteria.

Only one student received ratings in both the "more than satisfactory" and "unacceptable" categories.

Table 2 presents a distribution of ratings across the five criteria. As discussed briefly above, a total of 39, 90, and 31 ratings were assigned in the "more than satisfactory," "satisfactory," and "unacceptable" categories, respectively. However, it can also be seen that the distribution of ratings was very consistent across evaluation criteria.

Overall, it appears that faculty judged the large majority of psychology students to be at least competent writers. While the distribution of ratings was varied, over 80% were in the top two categories. Finally, there is a suggestion that individual criteria are not necessary to differentiate quality of student writing. With only one exception, individual student performance was judged in the same or adjacent categories on all criteria. The scale that was developed to assess student writing quality did not include a single overall evaluation for individual students. However, for each student, performance in any criterion category seemed to closely match that in other categories.

## *General Conclusions and Applications*

Assessment is valuable primarily as a guide to improving performance. As a result of this assessment of student writing, graduate teaching assistants (GTAs) who teach the research methods lab sections are now introduced to the criteria that were developed during assessment and instructed how to use these criteria in grading student writing. The assessment also informed faculty and GTAs of what writing skills need to be emphasized in teaching the research methods course. Grading is now more holistic and based less on minor editorial details than it was prior to the assessment. Furthermore, faculty, department-wide, have been encouraged to incorporate the criteria into the evaluation of their own students' writing.

Holistic writing assessment, due to its reliance on criteria developed for specific types of writing, can be successfully applied to writing in all disciplines and to different types of writing in a single discipline. Such assessment, if done well, requires extensive training and agreement by assessors of what criteria are important and how these criteria are to be operationalized. However, when implemented appropriately, holistically-derived criteria can provide a framework for instruction and evaluation of writing, regardless of its purpose or focus.

## References

Cooper, C. R. (1977). Holistic evaluation of writing. In Cooper, C. R. & Odell, L., *Evaluating Writing: Describing, Measuring, Judging* (p. 3-31). Urbana, IL: National Council of Teachers of English.

Faculty who participated in the writing assessment are Louis C. Buffardi, Linda D. Chrosniak, Carol Erdwins, Michael E. Hurley, Stefanie M. Joswick, Evans Mandes, James F. Sanford, and Merry J. Sleigh.

Table 1. Distribution of ratings by qualitative category. Interpretation: two papers were rated “more than satisfactory” on all five criteria, 1 paper was rated “more than satisfactory” on four criteria and “satisfactory” on one criterion.

Table 1.

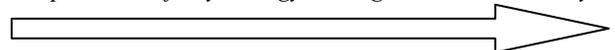
<b>Number of Papers</b>	<b>More than Satisfactory</b>	<b>Satisfactory</b>	<b>Unacceptable</b>
2	5		
1	4	1	
4	3	2	
4	2	3	
4	1	4	
1	1	3	1
1		5	
4		4	1
9		3	2
2		1	4
<b>32 Total Papers</b>			

Table 2.

<b>Number of Papers</b>				
<b>Criteria</b>	<b>More than Satisfactory</b>	<b>Satisfactory</b>	<b>Unacceptable</b>	<b>Total Papers</b>
Content	9	16	7	32
Flow	8	19	5	32
Format	8	16	8	32
Mechanic	6	19	7	32
Audience	8	20	4	32
<b>Total Rating</b>	<b>39</b>	<b>90</b>	<b>31</b>	

Notes: Interpretation: Nine papers were rated “more than satisfactory,” sixteen were rated “satisfactory,” and seven were rated “unacceptable” on the “content” criterion.

*Criteria/Scoring Sheet for Writing Assessment  
Department of Psychology, George Mason University*



**Attachment 1.**  
**Criteria/Scoring Sheet for Writing Assessment**  
**Department of Psychology**

**January, 2002**

1. Content is clearly stated, including adequate justification of hypotheses, appropriate level of detail throughout, and reasoned/logical presentation of research. Literature review studies are applicable and used appropriately, and the review includes adequate and appropriate documentation. Method section includes sufficient detail to enable reader to replicate the study.

More than satisfactory \_\_\_\_\_ Satisfactory \_\_\_\_\_ Unacceptable \_\_\_\_\_

2. The text “flows” well. Flow includes (a) organization (clear, well-developed hypotheses; clear topic sentences; good transition between ideas; all sections of paper tie together) and (b) tone (language is professionally appropriate; word choice is accurate; text demonstrates appropriate voice for writing task).

More than satisfactory \_\_\_\_\_ Satisfactory \_\_\_\_\_ Unacceptable \_\_\_\_\_

3. APA style and format are used correctly, including format of reference section and agreement between references cited in text and those in reference section.

More than satisfactory \_\_\_\_\_ Satisfactory \_\_\_\_\_ Unacceptable \_\_\_\_\_

4. Mechanics are correct. These include grammar, syntax, spelling, punctuation, and use of correct homonyms/near homonyms (e.g., affect, effect; compliment, complement) while avoiding non-word errors (e.g., alot).

More than satisfactory \_\_\_\_\_ Satisfactory \_\_\_\_\_ Unacceptable \_\_\_\_\_

5. Paper is written for the appropriate audience, namely, individuals who read research articles.

More than satisfactory \_\_\_\_\_ Satisfactory \_\_\_\_\_ Unacceptable \_\_\_\_\_

*This In Focus and earlier editions of this publication can be found at our website,  
<http://assessment.gmu.edu>*

---

**George Mason University**