
FACULTY FORUM

Developing a Rubric to Assess Student Learning Outcomes Using a Class Assignment

Nicholas Thaler, Ellie Kazemi, and Crystal Huscher
California State University, Northridge

We developed a rubric to assess several of our department's undergraduate student learning outcomes (SLOs). Target SLOs include applications of principles of research methodology, using appropriate statistics, adherence to the Publication Manual of the American Psychological Association, and written communication skills. We randomly sampled 20 percent (N = 55) of the final written manuscripts from several sections of a research methods course and trained 2 graduate-level raters to use the rubric to score the students' papers. We found statistically significant interrater reliability and convergent validity coefficients. These findings are discussed to encourage the development and evaluation of such rubrics to be used across colleges and universities.

In the last decade, undergraduate psychology programs have responded to the need for assessment as a legitimate force in higher education (see Dunn, Mehrotra, & Halonen, 2004). As such, there is an increasing interest in multimethod assessment of student learning outcomes and the development of reliable and valid rubrics to directly measure student learning (e.g., Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz, 2009). However, it is quite challenging to develop rubrics that reliably assess student learning based on oral and visual presentations, group work and discussions, final projects, written reports, and other culminated student work.

Halonen et al. (2003) created a rubric specifically designed for psychology departments to assess undergraduate students' scientific inquiry skills. Scientific inquiry is defined as forming hypotheses, de-

signing experiments, interpreting outcomes, and communicating results. These four domains converge with McGovern and Hawks's (1986) conclusions on what defines psychology student learning objectives (SLOs)—specifically that scientific reasoning and proficiencies in research methods and communication are their principal components.

The rubric that Halonen and colleagues (2003) designed is comprehensive and broad, incorporating elements of communication, collaboration, and self-assessment, as well as scientific inquiry skills. Each domain has five levels of proficiency, ranging from before training to professional graduate and beyond. They designed the rubric to measure a student's progress throughout his or her academic experience, from as early as high school up to college graduation.

One limitation Halonen and colleagues (2003) acknowledged is the time and energy required to effectively use a comprehensive multidimensional rubric. Consequently, some educators might lose enthusiasm in this rubric's implementation. This limitation can be addressed by focusing on a simpler rubric and using it as one part of a multimethod model of assessment. Furthermore, Halonen and colleagues' rubric was not empirically tested for reliability and validity. The authors hoped that their article would encourage other researchers to investigate the potential of developing their own rubrics to quantify and assess psychology education achievement. It is possible that designing a more specific rubric for a single assignment or final project might provide educators with an easier way of evaluating their students' achievement, provided that

the assignment or final project is representative and appropriately selected.

The purpose of this study was to follow Halonen et al.'s (2003) footsteps and to empirically develop a rubric that assesses achievement of certain learning outcomes using psychology students' research manuscripts. We constructed a rubric based on the criteria of the *Publication Manual of the American Psychological Association* (APA, 2001). This rubric was designed to specifically measure the degree to which our students achieved some of the SLOs of our department. Our rubric has a strong basis for its design in that most of its content was directly lifted from the *APA Manual*. We hope that our findings will demonstrate the potential of designing rubrics to assess psychology SLOs in both an objective and reliable manner.

Method

Participants

For this study, we randomly selected the names of 20% of the enrolled students in each section of a research methods course and sent the names of the selected students to the course instructors. All instructors offered copies of the selected students' final papers which resulted in a total of 55 manuscripts (women = 52, men = 3). Because the data contained no identifying information, the California State University, Northridge (CSUN) Institutional Review Board for Protection of Human Subjects approved this study.

Materials

The Psychology Department at CSUN has identified eight specific SLOs adopting the Task List goals established by the American Psychological Association Task Force (2007). CSUN's SLOs encompass many of the skills required to conceptualize and design an experiment, use the appropriate statistical tests, and write a manuscript in accordance with the APA guidelines. The rubric has 10 items that assess four of our SLOs: the use and interpretation of statistical techniques, critical thinking skills and skeptical inquiry in evaluating their own and others' research, competencies in electronic and information technologies, and effective written communication skills. Each of the 10 categories has a 6-point Likert scale. A copy of the rubric can be viewed at http://docs.google.com/Doc?id=df6b863n_0dw8dm3gj.

Procedure

A focus group of faculty members initially developed the rubric. We recruited two raters and revised the rubric twice based on the raters' feedback. Of the 55 manuscripts, both raters coded 22 (i.e., 40%) for assessment of interrater reliability. We also compared the final ratings of the manuscripts that were based on the rubric with each individual instructor's evaluation methods.

Results and Discussion

The interrater reliability between the two raters was analyzed using Spearman's correlations (see Table 1).

Table 1. Mean, Medians, Standard Deviation, Inter-Rater Reliability Between the Two Coders, and Convergent Validity

	M	Mdn	SD	IRR r_s	CV rp
1. Adherence to APA Style	3.9	4	1.0	.58**	.34
2. Quality of Abstract	4.2	5	1.1	.65**	.43
3. Relevant Literature	4.1	4	1.0	.59**	.74**
4. Hypotheses Oper. Defined	3.6	4	1.3	.35	.46
5. Quality of Methods	3.2	3	1.1	.69**	.51*
6. Quality of Results	3.4	3	1.4	.89**	.29
7. Quality of Discussion	3.8	4	1.0	.70**	.70**
8. Statement Pros/Cons	3.9	4	1.1	.64**	.23
9. Use of Charts/Figures ¹	2.8	3	1.5	.79**	-.18
10. Written Communication	3.5	4	1.1	.92**	.71**

*Note. $p < .05$

*Note. $p < .01$

¹Note. Data does not include missing charts and represents 91% of the data. (50 out of 55)M

Average interrater reliability was $r_s(22) = .68, p < .01$. The Quality of Results, $r_s(22) = .89, p < .01$, and the Overall Written Communication items, $r_s(22) = .92, p < .01$, had the strongest correlations, and the correlation coefficient of the Hypotheses Operationally Defined item was not statistically significant, $r_s(22) = .35, p < .11$. All other correlations between individual items ranged from .58 to .79 and were large (see Cohen, 1988, for interpretations of correlation coefficients as related to reliability).

We analyzed convergent validity using Pearson correlations (see Table 1). Each of the 10 scores assigned on the rubric as well as an overall average score was compared to the instructors' given scores of the papers from the course. Instructors used their own methods for grading the papers that were unrelated to this developed rubric. Not all instructors were available to provide their past scores of the individual assignments. Results show that overall the rubric correlated with actual scores with a mean of $r(17) = .56, p < .01$. Individual items varied, but the Quality of Introduction item, $r(17) = .74, p < .01$, the Quality of Methods item, $r(17) = .51, p < .05$, the Quality of Discussion item, $r(17) = .70, p < .01$, and the Written Communication Skills item, $r(17) = .71, p < .01$, all yielded significant correlations. All other individual item correlations were not significant.

We constructed this assessment rubric under the guidelines of previous papers on rubric development and standardization. Although it has only been tested on one particular sample, it appears to be an effective measure for assessing certain SLOs of the CSUN undergraduate Psychology Department. More specifically, the rubric can be used as a direct assessment of written communication skills, use of appropriate statistics, use of previous literature to support ideas and hypotheses, adherence to the *APA Manual*, and ability to discuss key concepts of research methodology.

The strong interrater reliability suggests that the meaning of each item was adequately transmitted to the raters, although there was a great deal of difficulty in explaining the qualities that a strong manuscript's statements of hypotheses entail. This item's low correlations can be attributed to the relatively obscure nature of the item itself, as guidelines were not available in the *APA Manual*. All other items have moderate to strong correlations between the two coders.

We found that items that depended strongly on writing ability were most similar to the actual grade of the manuscript, and items that focused on statistics and research methodologies were dissimilar. It is possible that professors valued writing ability above all other SLOs

when judging the manuscripts due to a halo effect. It is also possible that the rubric has difficulty in capturing the professors' interpretations of the students' research methodologies and knowledge of statistics.

Limitations and Future Research

One limitation of our study is a lack of intrarater reliability, as the raters only coded each manuscript once. Another limitation is the relatively small sample of papers analyzed for convergent validity, as some faculty members were unavailable to provide permission in accessing their actual final paper assigned scores.

The possible halo effect of a strongly written paper might have interfered with the true quality of the manuscripts. It is difficult to ascertain whether a manuscript received a high score for demonstrating mastery of our SLOs, or merely for being well written. Conversely, it could be that instructors, raters, or both underestimated a poorly written paper and consequently assigned it lower scores in other content areas.

This project provided an example of constructing an item-specific rubric with explicit guidelines on what is expected on each section of a psychology manuscript. Future studies should describe the process and challenges of other researchers in developing their own assessment rubrics for SLOs. Although we designed the rubric for use in one psychology department, we hope that these results will provide an addition to the relatively sparse information on the empirical development of reliable and valid rubrics designed to measure undergraduate psychology students' academic performances.

References

- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association Task Force on Psychology Major Competencies. (2007). *APA guidelines for the undergraduate psychology major*. Washington, DC: Author.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Dunn, D. S., Mehrotra, C. M., & Halonen, J. S. (Eds.). (2004). *Measuring up: Educational assessment challenges and practices for psychology*. Washington, DC: American Psychological Association.
- Halonen, J. S., Bosack, T., Clay, S., & McCarthy, M. (with Dunn, D. S., Hill, G. W., IV, McEntarffer, R., Mehrotra,

- C., Nesmith, R., Weaver, K. A., & Whitlock, K.). (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30, 196–208.
- McGovern, T. V., & Hawks, B. K. (1986). The varieties of undergraduate experience. *Teaching of Psychology*, 31, 174–181.
- Stellmack, 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*.

Notes

1. We thank Dr. Shannon Morgan for her detailed and helpful comments on an earlier draft of this article. We also thank the CSUN Psychology Department faculty for their participation, input, and support.
2. Send correspondence to Ellie Kazemi, Department of Psychology, California State University, Northridge, CA 91330–8255; e-mail: ellie.kazemi@csun.edu.