Development and Validation of a Universal Science Writing Rubric That is Applicable to Diverse Genres of Science Writing

Alycia Pisano,a Amanda Crawford,b Heather Huffman,c Barbara Graham,c and Nicole Kelp,c

aDepartment of Chemical and Biological Engineering, Colorado State University, Fort Collins, Colorado, USA
bColorado School of Public Health, Aurora, Colorado, USA
cDepartment of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado, USA

KEYWORDS rubric, writing, science communication, assessment tool

INTRODUCTION

Science, technology, engineering, and mathematics (STEM) students must develop communication skills, including science writing, to become effective future scientists. Thus, STEM curricula often involve writing, including lab reports, for students to develop these skills. Many instructors follow the best practice of providing a rubric when grading student writing (1). However, these rubrics are created with various goals for each assignment, so a “90%” on a paper graded with one rubric is not comparable to a “90%” on a paper graded with a different rubric, precluding programmatic assessment. Additionally, instructors and teaching assistants often grade more on surface features like grammar instead of discourse and content (2, 3).

Some published science writing rubrics, including the “Conclusion Assessment Rubric” (4) and “Rubric for Science Writing” (5), can provide more consistency, but these are focused on laboratory/data reports and are not translatable for other genres. In addition to traditional “science writing” like lab reports or research papers, assessing students’ skills in communicating science to nonscientific audiences is important. A published instrument for assessing science writing geared toward public audiences focuses on analogy, narrative, and dialogue (6), making it appropriate for lay writing but less applicable to science journal articles.

We thus created a Universal Science Writing Rubric (USWR) that can apply to all genres of science writing from the classroom and laboratory, regardless of intended audience. Our rubric was informed by a multidisciplinary perspective (7, 8), bringing theory and methods from science education (9, 10), science communication (11, 12), and applied linguistics (13, 14). As shown in Table 1, the USWR is a tool to assess various rhetorical concerns in students’ writing, including clarity of scientific content, interpretation of scientific content, targeting the audience, organization, and writing quality. In this article we present the USWR, implementation ideas, and sample data that it can produce to demonstrate its validity and practicality for diverse science writing genres.

PROCEDURE

The USWR is a useful tool for student self-assessment, peer grading (15), instructor or teaching assistant grading (3), and departmental programmatic assessment of any science writing from the classroom or laboratory. The USWR is also useful for science education researchers to complete pre- and post-assessment of curricular interventions regarding science writing. Because the USWR does not simply check for the inclusion of assignment-specific components but instead focuses on competencies that cross genres, the USWR enables students, teachers, and researchers to holistically assess the progression of students’ skills over time.

The USWR could be modified in many ways to support diverse needs as follows.

- In our preliminary testing, we assigned the 4 levels of scores of 0 to 3. These numbers could be increased to recognize student effort.
- The rubric could be used honestly across students in a department, but first year students only need to score 5/15 to earn 100% while higher-level students need a 12/15 to earn 100%. This prevents students from encountering less precise rubrics giving them high scores freshman year only to be blindsided by lower grades as rubrics become
| Parameter                        | Absent                                                                 | Emerging                                                                                                                                   | Proficient                                                                                                                                                                                                 | Mastery                                                                                                                                                                                                 |
|---------------------------------|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Scientific content (C)          | Scientific content presented is inaccurate                            | Scientific content presented is accurate, but there are elements of the scientific story missing (either scientific findings or process are described insufficiently) | Scientific content presented is accurate, and both findings and process are described, but the story may be disjointed                                                                                       | Scientific content (both findings and process) is accurate, and scientific ideas are integrated to tell a story                                                                                           |
| Interpretation of scientific content (I) | There is no interpretation of the scientific findings OR there was an incorrect interpretation (e.g., false confidence, correlations presented as causations) | An attempt was made to interpret the scientific findings and place them in the context of the field; scientific uncertainty or limitations are mentioned | There is a deeper discussion that interprets the implications and/or limitations of the studies in the context of the field                                                                                     | The interpretation is holistic, discussing implications and uncertainties of the findings in the context of the field, and is explained well for the genre of the paper                                           |
| Targeting the audience (T)      | The writing was not targeted well to the intended audience (e.g., the main thesis of the writing was not appropriate for the intended audience) | An attempt was made to gear the writing toward the intended audience, but there were still issues with the level of detail (too detailed content or not enough description) for the audience | An attempt was made to gear the writing toward the intended audience, but there were still small issues with language (e.g., too much jargon for a lay audience, too colloquial for a scientific audience) | Content, organization, and language were all geared appropriately towards the intended audience                                                                                                          |
| Organization (O)                | The flow of information is in an illogical order                       | The order of information is mostly logical and/or may not fit the genre/goals of the paper well                                           | The order of information presented is logical for the genre, but smooth transitions are lacking                                                                                                             | The order of information presented is logical for the genre, and transitions between concepts are smooth                                                                                                   |
| Writing quality (W)             | The writing is mostly inaccurate in terms of accepted grammatical conventions/sentence structure/word choice | The writing is readable but has noticeable errors or issues with grammatical conventions/sentence structure/word choice | The writing is mostly accurate in terms of accepted grammatical conventions/sentence structure/word choice for the genre, with only minor errors                                                               | The writing is accurate in terms of accepted grammatical conventions/sentence structure/word choice for the genre and is smooth to read                                                                   |
more discriminating. Students would see needed areas for improvement as younger students, without this penalizing their actual grade.

- Each rubric category could be weighted differently to support the hierarchy of rhetorical concerns (16), with content, targeting, and interpretation being weighted higher than organization and writing quality.

## CONCLUSION

### Preliminary testing

Student writing samples (Table 2) were collected in life science departments at two universities (Washington State University institutional review board [IRB] no. [18121-001] and Colorado State University IRB no. 20-10236H). Two coders scored each writing sample and then discussed until interrater reliability was achieved, as indicated by an intraclass correlation coefficient of >0.8 (17–19). We provide examples of what constituted each rubric score (see Appendix S1 in the supplemental material) and frequently asked questions (FAQs) about use of the USWR (see Appendix S2 in the supplemental material). The nonparametric Wilcoxon rank test or signed-rank test for paired samples (20, 21) was used to compare sets of student writing samples. The Kruskal-Wallis test (22) was used for testing three levels of writing, followed by Dunn's test (23); a P value of <0.05, adjusted for multiple comparisons, indicated significance. All statistics were calculated using R (packages in Appendix S3 in the supplemental material).

The USWR revealed the following statistically significant trends in student science writing.

1. **Draft versus final versions of diverse genres of writing.** Overall scoring was higher on final versions than draft versions. However, some individual students did not improve from draft to final or even received lower final scores, confirming that students likely need further training in responding to instructor or peer feedback (24).

2. **Papers of the same genre at different levels.** In our samples, students' scientific interpretation and targeting skills increased from 200- to 400-level but did not change from 400- to 600-level (Fig. 1A). The higher-level samples showcased outliers; this supports observations that as students gain independence in graduate school, they still have room for improvement in science writing (25).

3. **Papers at the same level but written to scientific versus lay audiences.** We found that students struggled to interpret science for lay audiences but overall struggled to properly target a scientific audience (Fig. 1B). Of note, the rubrics for the original paper did not include targeting, suggesting that students write to the rubric rather than considering all aspects of quality science writing.
4. Lab reports submitted by students at the beginning, middle, and end of a semester. There was no significant change in scores over time. This demonstrates that a consistent rubric could facilitate the development of writing skills. The teaching assistant grading these lab reports indicated that grading rubrics were based on inclusion of particular items in each individual lab report and were not consistent from report to report, validating the data produced by the USWR.

We also compared the USWR to grading rubrics used for several of the writing assignments collected in this study. The USWR was more discriminating (producing lower average scores and higher standard deviations) than the grading rubrics, pinpointing students who were struggling with particular skills but may have received a passing grade with a traditional rubric.

Discussion

The USWR is an important tool for assessing student writing skills. While our example data is limited by small $n$ values, we show that the USWR is usable for diverse science writing genres, sensitive enough to reveal statistically significant differences between writing samples, and valid enough to produce findings supported by the literature and instructor observations. Overall, benefits of using our USWR include better discrimination than some grading rubrics but discrimination based on important rhetorical concerns like scientific interpretation and targeting rather than surface-level grammatical features.

FIG 1. Example data produced by USWR analysis of student science writing, highlighting statistically significant differences in interpretation and targeting. (A) Analysis of writing to a scientific audience at different levels. Students improved from the 200 level to the 400 level, but there was no difference between the 400 and 600 level. Gray bar indicates median. (B) Analysis of writing to a scientific versus a lay audience at the same level. Lines connect paired scores (from the same student). Students were better at interpreting for a scientific audience but better at targeting for a lay audience. For both panels A and B, there were no statistically significant differences between sets of samples in other rubric categories.
ACKNOWLEDGMENTS

This work was funded by startup funds provided to Nicole Kelp by Colorado State University.

We thank the instructors at two universities who facilitated collection of writing samples from students in their courses.

We declare no conflicts of interest.

REFERENCES

1. Clabough EBD, Clabough SW. 2016. Using rubrics as a scientific writing instructional method in early stage undergraduate neuroscience study. J Undergrad Neurosci Educ 15: A85–A93.
2. Szymanski EA. 2014. Instructor feedback in upper-division biology courses: moving from spelling and syntax to scientific discourse. Across the Disciplines 11:1–14. https://doi.org/10.37514/ATD-J2014.11.206.
3. Hill CFC, Gouvea JS, Hammer D. 2018. Teaching assistant attention and responsiveness to student reasoning in written work. CBE Life Sci Educ 17:ar25. https://doi.org/10.1187/cbe.17-04-0070.
4. Cary T, Harris M, Hong S, Yin Y. 2019. Conclusion assessment rubric (CAR). Society for the Advancement of Biology Education Research, Minneapolis, MN.
5. Timmerman BEC, Strickland DC, Johnson RL, Payne JR. 2011. Development of a ‘universal’ rubric for assessing undergraduates’ scientific reasoning skills using scientific writing. Assess Eval Higher Ed 36:509–547. https://doi.org/10.1080/02602930903540991.
6. Baram-Tsabari A, Lewenstein BV. 2013. An instrument for assessing scientists’ written skills in public communication of science. Sci Commun 35:56–85. https://doi.org/10.1177/1075547012440634.
7. Stoller F, Robinson M. 2014. Drawing upon applied linguistics to attain goals in an interdisciplinary chemistry-applied linguistics project, p 11–25. In Curry MJ, Hanauer DI (ed), Language, literacy, and learning in STEM education: research methods and perspectives from applied linguistics. John Benjamins Publishing Company, Amsterdam, Netherlands.
8. Murdock RC. 2017. An instrument for assessing the public communication of scientists. PhD dissertation. Iowa State University, Digital Repository, Ames, IA.
9. Reynolds JA, Thaiss C, Katkin W, Thompson RJ. 2012. Writing-to-learn in undergraduate science education: a community-based, conceptually driven approach. CBE Life Sci Educ 11:17–25. https://doi.org/10.1187/cbe.11-08-0064.
10. Hand B, Wallace CW, Yang E-M. 2004. Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. Int J Sci Educ 26:131–149. https://doi.org/10.1007/0950069032000070252.
11. Nisbet MC, Scheufele DA. 2009. What’s next for science communication? Promising directions and lingering distractions. Am J Bot 96:1767–1778. https://doi.org/10.3732/ajb.0900041.
12. Fischhoff B, Davis AL. 2014. Communicating scientific uncertainty. Proc Natl Acad Sci USA 111(Suppl):13664–13671. https://doi.org/10.1073/pnas.1317504111.
13. Hanauer D, Curry MJ. 2014. Integrating applied linguistics and literacies within STEM education: Studies, aims theories, methods, and forms, p 1–8. In Language, literacy, and learning in STEM education: research methods and perspectives from applied linguistics. John Benjamins Publishing Company, Amsterdam, Netherlands.
14. McCarthy M. 2001. Applying linguistics: disciplines, theories, models, descriptions. In Issues in applied linguistics. Cambridge University Press, Cambridge, United Kingdom.
15. Deng Y, Kelly G, Deng S. 2019. The influences of integrating reading, peer evaluation, and discussion on undergraduate students’ scientific writing. Int J Sci Educ 41:1408–1433. https://doi.org/10.1080/09500693.2019.1610811.
16. Colorado State University Writing Center. Hierarchy of rhetorical concerns. Colorado State University, Fort Collins, CO.
17. Bartko JJ. 1966. The intraclass correlation coefficient as a measure of reliability. Psychol Rep 19:3–11. https://doi.org/10.2466/pr0.1966.19.1.3.
18. Shroft PE, Fleiss JL. 1979. Intraclass correlations: uses in assessing rater reliability. Psych Bull 86:420–428. https://doi.org/10.1037/0033-2909.86.2.420.
19. Shroft PE, Fleiss JL, Wright JT. 1981. Intraclass correlations: a guide for use in the evaluation of measurement error. Psych Bull 88:524–535. https://doi.org/10.1037/0033-2909.88.4.524.
20. Whitley E, Ball J. 2002. Statistics review 6: nonparametric methods. Crit Care 6:509–513. https://doi.org/10.1186/cc1820.
21. Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics Bull 1:80–83. https://doi.org/10.2307/3001968.
22. Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621. https://doi.org/10.1080/01621459.1952.10483441.
23. Dunn OJ. 1964. Multiple comparisons using rank sums. Technometrics 6:241–252. https://doi.org/10.1080/00401706.1964.10490181.
24. Ornerella Treglia M. 2008. Feedback on feedback: exploring student responses to teachers’ written commentary. J Basic Writing 27:105–137. https://doi.org/10.3751/JBWr.J2008.27.1.06.
25. Wagenmakers E-J. 2009. Teaching graduate students how to write clearly. APS Observer 22.