

Rubric Development for Technical Reports in Chemical Engineering Unit Operations Laboratory Courses

Dr. Jennifer R. Brown, Montana State University, Bozeman

Jennifer Brown is an Associate Professor in the Chemical and Biological Engineering Department at Montana State University in Bozeman MT.

Prof. Stephanie G. Wettstein, Montana State University, Bozeman

Stephanie Wettstein is an Assistant Professor in the Chemical and Biological Engineering department at Montana State University in Bozeman, MT. She has been the faculty advisor of the MSU SWE chapter since 2013.

Dr. Douglas J. Hacker, University of Utah

Dr. Hacker is Professor Emeritus in the Department of Educational Psychology and participates in both the Learning Sciences Program and the Reading and Literacy Program.

Rubric development for technical reports in chemical engineering unit operations laboratory courses

Abstract

The purpose of this work was to test the inter-rater reliability (IRR) of a rubric used to grade technical reports in a senior-level chemical engineering laboratory course that has multiple instructors that grade deliverables. The rubric consisted of fifteen constructs that provided students detailed guidance on instructor expectations with respect to the report sections, formatting and technical writing aspects such as audience, context and purpose. Four student reports from previous years were scored using the rubric, and IRR was assessed using a two-way mixed, consistency, average-measures intraclass correlation (ICC) for each construct. Then, the instructors met as a group to discuss their scoring and reasoning. Multiple revisions were made to the rubric based on instructor feedback and constructs rated by ICC as poor. When fair or poor constructs were combined, the ICCs improved. In addition, the overall score construct continued to be rated as excellent, indicating that while different instructors may have variation at the individual construct level, they evaluate the overall quality of the report consistently. A key learning from this process was the importance of the instructor discussion around their reasoning for the scores and the importance of an 'instructor orientation' involving discussion and practice using the rubrics in the case of multiple instructors or a change in instructors. The developed rubric has the potential for broad applicability to engineering laboratory courses with technical writing components and could be adapted for alternative styles of technical writing genre.

Introduction

Technical communication skills are highly valued in industry[1, 2] and tied to higher levels of career advancement[3]. Regrettably, they are often lacking in engineering graduates. In chemical engineering programs without a technical writing course requirement, technical communication instruction is often incorporated into the curriculum through laboratory-based or capstone design courses[4-7]. In such courses, designing meaningful communication assignments and assessments is often relegated to engineering faculty, who may or may not have the requisite skills and knowledge to do so.

Student learning of communication skills is tied to quality of feedback[4]; however, engineering faculty typically have not received any formal training on how to effectively give feedback on technical writing[2, 4]. Engineering faculty may tend to focus on spelling and grammar, while effective feedback is higher level and corresponds to issues with organization, the use of arguments, or support of evidence[4, 8]. Good feedback is essentially more of a coaching rather than correcting[2], and collaboration with communication experts for training is one approach to develop more efficient and purposeful grading rubrics. The goal of rubrics is to reflect the skills targeted in the assignment in order to effectively evaluate technical communication[2].

In previous work, the lead instructors for a two-part series of senior-level chemical engineering unit operations laboratory courses worked with the Writing Center on campus to develop assignments and activities targeted at specific technical communication skills[9]. Through this collaboration, preliminary rubrics were developed to assess communication skills tied to learning outcomes. These rubrics were constructed based on reflecting what key information the students were to convey and course objectives. Additionally, considerable thought went into what would

cause the students to not meet expectations and lose points for each of the constructs in the rubrics. Well-designed rubrics can help faculty set clear expectations for students, provide feedback and assess technical writing skills[10]. Additionally, it is important for rubrics to be reliable across instructors in team taught courses or when instructors change.

This study aimed to evaluate the inter-rater reliability (IRR) of the technical report rubric developed in collaboration with the Writing Center across instructors teaching laboratory courses within the chemical engineering curriculum. We present the results of this evaluation as well as lessons learned from the scoring discussion. Additionally, we provide recommendations for incorporating an ‘instructor orientation’ prior to using rubrics to ensure effective use of the rubric across multiple instructors.

Methods

The rubric underwent two rounds of validation. The first was in the 2022 spring semester and the second followed in the 2022 fall semester. For the spring rubric validation, six instructors graded four reports and used a rubric that consisted of 15 constructs ranging in value from 4 to 12 points each (Table 1) for a total of 100 points. Each construct on the rubric had five levels meant to correspond to “A”, “B”, “C”, “D” and “F” level work. The description of “A” level from the rubric is listed in Table 1.

In the case of the fall rubric, three constructs were removed and two new constructs were added (Table 1), Safety (“A” level description: “*Contains all key safety concerns, hazards, and how each issue will be handled considering high probability and most likely outcomes*”) and Overall Report (“A” level description: “*Work earning this score is ready to be passed on to a real client. In every way, it meets audience needs. Document/presentation is formatted and organized to guide to major points. Clear and interesting visuals and prose contribute to professional-level quality. Overall report considerations include: Audience – Demonstrates a thorough understanding of the audience and purpose that is responsive to the assigned task and who the report was to be written for; Tech Writing – Well-ordered, segued, logical flow of material; accurate spelling, punctuation, grammar and sentence structure; clear and concise; correct amount of detail; own words used; good transitions between sentences, paragraphs, and sections; Format – Used 11 pt font minimum and an acceptable font; equation editor was used for equations and for variables in main text; no large white spaces; formatting was uniform throughout document; correct sig figs; equations inline; Figure and Table Formatting – Concise, but descriptive captions that include symbols if needed are present, correct location, and are referred to in the text; figures have axis labels, no title, no outer border, no legend, black font and lines, no extra decimal places; tables are in a simple layout format*”). For the fall rubric, the constructs ranged in value from 4 to 24 points for a report total of 100 points. Four new reports were graded by four instructors.

Inter-rater reliability (IRR) was assessed using a two-way mixed, consistency, average-measures intraclass correlation (ICC) for each of the constructs on which students were rated. The ICC is a descriptive statistic used to assess the level of consistency in ratings from two or more raters on the same construct across participants. The ICC works well with multiple raters who have used ordinal data for their ratings. The ICC can range from 0, random agreement, to +1.0, perfect agreement, with higher ratings indicating higher consistency among raters. For example, a rating

of 0.80 would indicate that 80% of the variance among raters was due to true consistency among raters, and 20% was due to unexplained variability or error. In general, ICCs less than 0.40 are poor, ICCs between 0.40 and 0.59 are fair, ICCs between 0.60 and 0.74 are good, and ICCs between 0.75 and 1.0 are excellent. Some caution must be used in interpreting the ICCs reported here mainly because only four reports were rated during each validation. However, the ICCs do provide a general sense of how consistent the raters were within a category.

Table 1: Constructs for the initial rubric (spring validation) and the revised rubric (fall validation) that were used for grading students' technical reports.

Spring Validation			Fall Validation	
Construct	Max Points	Description for "A" work	Construct	Max Points
1) Cover page	4	Completely descriptive title; authors/group# listed; date present	1) Cover page	4
2) Abstract	8	Excellent background and org/tech needs; quantitative results with ranges, error, uncertainty; definite conclusions on all key findings; excellent recommendations	2) Abstract	8
3) Intro & background	8	Completely relevant background, covering organizational, technical needs, and basic approach very well, yet concisely. Also includes brief, necessary theory (eqns) for audience to understand lab	3) Intro & background	8
4) Objectives	4	The objectives present a very clear picture of the experimental and stats goal of the lab. The objectives utilize active verbs that can be measured and that are not redundant	4) Objectives	4
5) Method	4	Contains all relevant sections, in particular a clear plan of work to accomplish all major objectives; detailed approach to deal with uncertainty; unnecessary procedural details omitted	5) Method	8
6) Results	8	All useful intermediate, final results; apropos chart and table use; excellent text to explain how results were obtained from data; accompanying uncertainty; references to eqns, methods, and critical assumptions or approx.. made in obtaining result	6) Safety	4
7) Discussion	8	All needed findings discussed in light of theory, work of others; <i>whys</i> given for all results; no significant conceptual errors; logical recommendations tied to theory and results	7) Results	8
8) Conclusions & Recomm	8	Purpose/objective is restated in a new way; major finding is clearly stated; revisits hypothesis; suggest good improvement/change to lab procedure and extension of lab and a recommendation to company	8) Discussion	8
9) References	4	At least four appropriate references used from reputable sources, cited properly in body, listed in separate section at end of report with clear/consistent format	9) Conclusions & Recomm	8
10) Appendices	4	Included: Completed data collection sheet, sample calculations, complete Excel analysis tables, safety section, step-by-step experimental protocol	10) References	4
11) Audience & purpose	8	Demonstrates a thorough understanding of context, audience, and purpose that is responsive to the assigned task(s) and who the report was to be written for (i.e., assigned task)	11) Appendices	4
12) Technical Writing	12	Well-ordered, segued, logical flow of material; accurate spelling, punctuation, grammar and sentence structure; clear and concise; own words used	12) Page Maximum	8
13) Format	8	Used 11 pt font minimum and an acceptable font; equation editor was used for equations and for variables in main text; no large white spaces; formatting was uniform throughout document; correct sig figs; eqn inline	13) Overall Report	24
14) Figure & Table Format	4	Concise, but descriptive captions that include symbols if needed are present, correct location, and are referred to in the text; figures have axis labels, no title, no outer border, no legend, black font and lines, no extra decimal places; tables are in a simple layout format		
15) Page Maximum	8	Report is less than 8 pages with at least 1 inch margins and 1.5 line spacing, which includes figures, tables, etc. (Page limit does not include pages of the appendix or references)		

All reports were from a senior-level chemical engineering laboratory course, and it should be emphasized that the spring and fall validations were done with different sets of student reports. That is, in the spring, four reports were selected from a laboratory experiment on heat exchangers, and for the fall, four reports from a different heat transfer modelling experiment were selected. For the spring validation, the six graders consisted of five graders who had been instructors of a laboratory course previously while one was a new teaching assistant for the class and had not graded technical reports previously. For the fall validation, three of the graders remained the same as in the spring validation and had all previously graded technical reports while one was a new instructor for the laboratory course and had not.

Results and Discussion

In the spring of 2022, the six graders evaluated four student reports on a heat exchanger experiment done in a senior-level laboratory class using the 15 constructs listed in Table 1. The point values for each of the 15 constructs were added together, which resulted in an overall report score that can be seen in Figure 1. Each colored line represents a different grader (see legend in table) and each corner labeled 1-4 represents a different report. Using Figure 1, discrepancies in the overall score of the report can be seen with most graders being consistent excluding the gray (high) and navy (low) graders for reports 2 and 4, and the teal (low) grader for report 3. When looking at the normalized point scores for each rubric construct for reports 2 and 4 (Figure 2), the amount of variability between instructors can be seen. The navy grader from Figure 1 (shown by navy diamond in Figure 2) had lower scores compared to the average across most rubric items.

The navy grader was a teaching assistant that had not graded reports previously and upon group discussion, it became apparent that it is important to discuss expectations and to review report grades with new graders to ensure consistency. Typically, as training for this laboratory course, an experienced instructor will grade two submitted student reports also graded by teaching assistants and other new instructors. They then meet to discuss the scoring and where there are discrepancies. If the scores are off by a substantial amount, more of the reports are scored in a similar fashion until agreement has been reached. Additionally, for the actual course, grades are

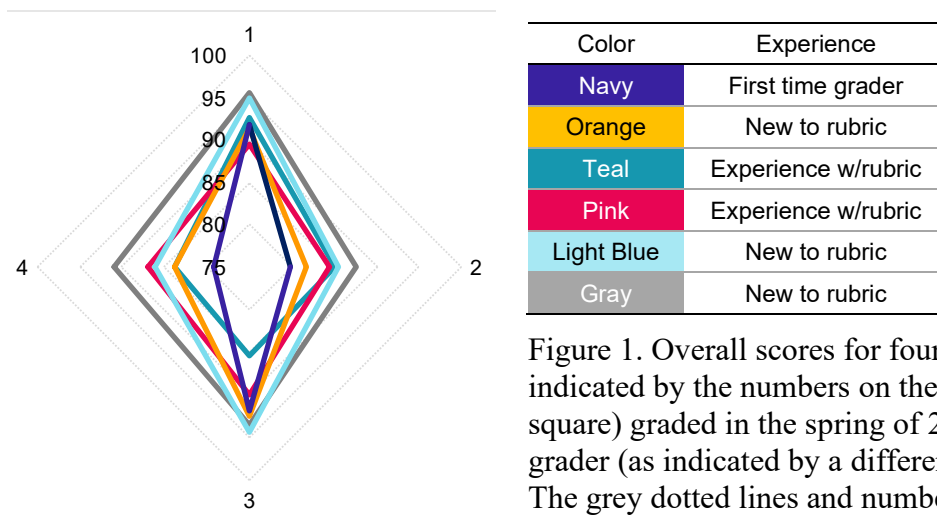


Figure 1. Overall scores for four reports (as indicated by the numbers on the points of the square) graded in the spring of 2022 for each grader (as indicated by a different colored line). The grey dotted lines and numbers are the axis values and correspond to the report scores.

entered by the new grader into the grading software but not published until the experienced instructor reviews the scoring to ensure it is within error of other graders. For the spring validation, the navy grader did not have this training prior to scoring the reports, which emphasizes a need to incorporate a validation session about expectations prior to using a rubric.

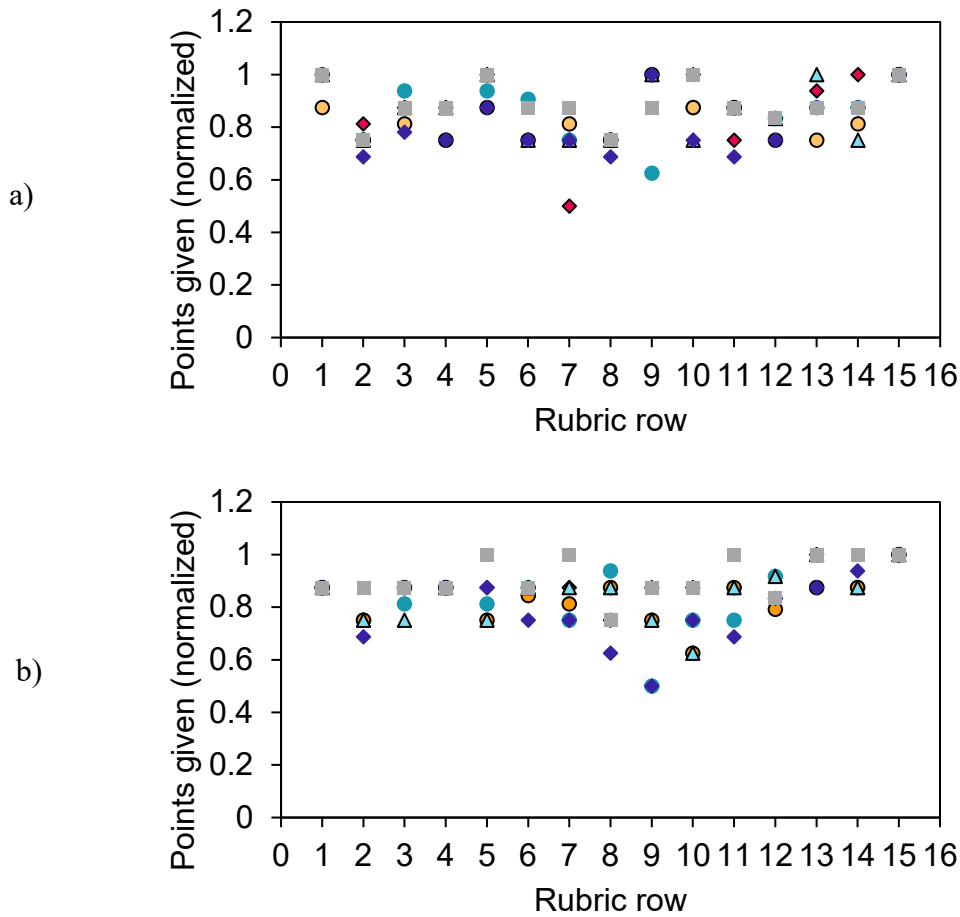


Figure 2: Scores from each of the six graders for student reports normalized by the maximum number of points in a construct for a) report 2 and b) report 4 from the spring rubric validation. The rubric row numbers on the x-axis correspond to the 15 rubric constructs (see Table 1) and the symbol colors correspond to grader colors listed in Figure 1.

As mentioned, following scoring of the reports, the instructors met to discuss the rubric and the scoring. For example, the teal grader identified statistical errors in report 3 that the other graders overlooked which was the reason for their lower grade. In addition, the gray grader had graded technical reports for the class in previous years but was not aware that fractional points could be given for a particular level in the new rubric. Consequently, they erred on the positive side when assigning point levels for a construct and gave the higher points, leading to the higher scores. Finally, as previously mentioned, the navy grader, a graduate-level teaching assistant, had not graded technical reports using a rubric before and had different expectations for the report

contents, resulting in different point values. While discussing the scores and rubric, several items were noted:

- The graders liked how the constructs were descriptive and allowed them to make assessments as they read through the report. For example, the constructs move from cover page to abstract to introduction, etc. The rubric was efficient to grade with because of this.
- The rubric was detailed and directed, which could benefit struggling students who need more guidance on what is expected.
- A holistic aspect was missing.
- Point values did not reflect the corresponding letter grade that was listed on the rubric. For example, for a rubric construct worth a maximum of 4 points, “A” work was labeled as 4 points while “B” work was labeled as 3 points and “C” work 2 points. However, if a student earned the points labeled as "C" work in all the constructs, they would receive a total score of 54 out of 100, which is an “F” grade rather than a “C” grade. Instructors felt the points labelled on the levels should be re-evaluated to more accurately reflect the letter grades.
- There were some parts of the rubric that double-counted points so discussion revolved around how to modify the constructs to minimize this from occurring. For example, if the student did not have a good discussion section, they may have lost points in that construct as well as in the technical writing construct.

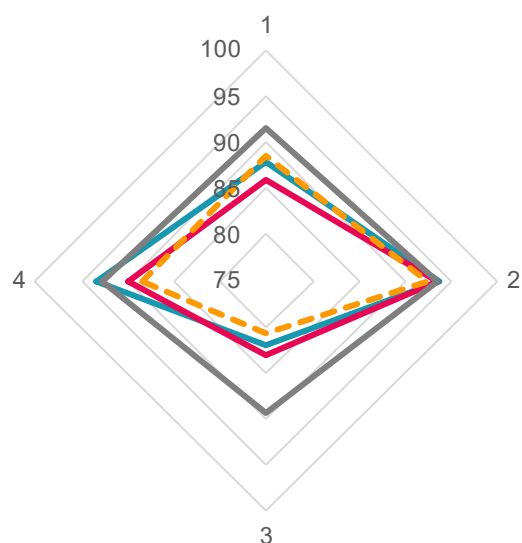
When ICCs were computed on the four reports graded by the six instructors, the results showed that overall, the raters provided fairly consistent ratings (Table 2). Seven of the constructs were rated as excellent, two as good, two as fair, and three as poor. In addition, the rating on Page Maximum was perfect, which is a “yes/no” category that gives full points if the report is under 8 pages or 0 points if it is over.

Table 2: The results of the spring validation IRR analysis.

Category	ICC Value	Rating
Cover page	0.716	good
Abstract	0.950	excellent
Intro/background	0.773	excellent
Objectives	0.600	good
Method	0.559	fair
Results	0.839	excellent
Discussion	0.861	excellent
Conclusions & Recomm	0.816	excellent
References	0.902	excellent
Appendices	0.921	excellent
Audience/purpose	0.447	fair
Technical Writing	0.194	poor
Format	0.058	poor
Figure/Table Format	0.138	poor
Page Maximum	perfect*	excellent
Total score	0.950	excellent

In order to take into account feedback from the discussion and address the constructs with poor ICCs, revisions were made to the rubric. The Technical Writing, Format, and Figures/Tables constructs were rated as ‘poor’ by the ICCs (Table 2), and these were the constructs that graders had concerns about double-counting errors. These three constructs were combined into a single holistic construct adapted from Sheffield et al.[11]. In addition, point values were adjusted to represent the letter grade earned and the resulting constructs and maximum point values are listed in Table 1.

In the fall of 2022, a new round of validation was conducted using the revised rubric. Four reports for a different lab experiment were scored by three of the previous rubric validators (shown by the same colors in Figures 1 and 3) and a new grader (indicated by the dashed line in Figure 3).



Color	Experience	Figure 1
Orange	New to rubric	New
Teal	Experience w/rubric	Same
Pink	Experience w/rubric	Same
Gray	Some exp. w/rubric	Same

Figure 3. Overall scores for four student reports represented by the four corners (1-4) with each color representing a grader. Scoring was completed in the fall of 2022 with the dashed orange line being a new grader compared to spring of 2022. The grey dotted lines correspond to report scores.

The scores showed less variability between three of the graders (Figure 3), but the gray grader continued to score high, particularly on reports 1, 3, and 4. The teal grader also gave a high score for report 4 compared to pink and orange. For reports 1 and 3, the score of the gray grader was outside one standard deviation of the average report score, which could be a concern. However, the teal score for report 4 was within the standard deviation for all graders (91.2% \pm 2.3%; Table 3) and was therefore not considered an issue. The average standard deviation of scores decreased from 3.1 using the original rubric to 2.3 using the revised rubric.

Table 3: Average scores and standard deviations of technical reports from spring (S) and fall (F) semester rubric validations.

Report	Ave Score \pm Std dev
S1	92.7 \pm 2.3
S2	84.0 \pm 2.8
S3	91.3 \pm 3.2
S4	85.1 \pm 3.9
F1	88.4 \pm 2.3
F2	93.3 \pm 0.6
F3	83.8 \pm 3.8
F4	91.2 \pm 2.3

When looking at discrepancies within each construct, for report 1 (Figure 4a) the gray grader gave more points in five of the constructs and less in two of the constructs. The largest discrepancy was in construct 8, “Discussion”, where they rated the paper an 8 compared to the 6.5, 5, and 6 given by the other graders. For report 3 (Figure 4b), the gray grader had less of a drastic difference, but was on the high-end of scoring for the majority of constructs. Since there are 13 constructs within the rubric, even an extra half point per construct will result in a 6.5% difference in final grade score. In this case, the gray grader’s score was approximately 7% higher than the average of the three other graders.

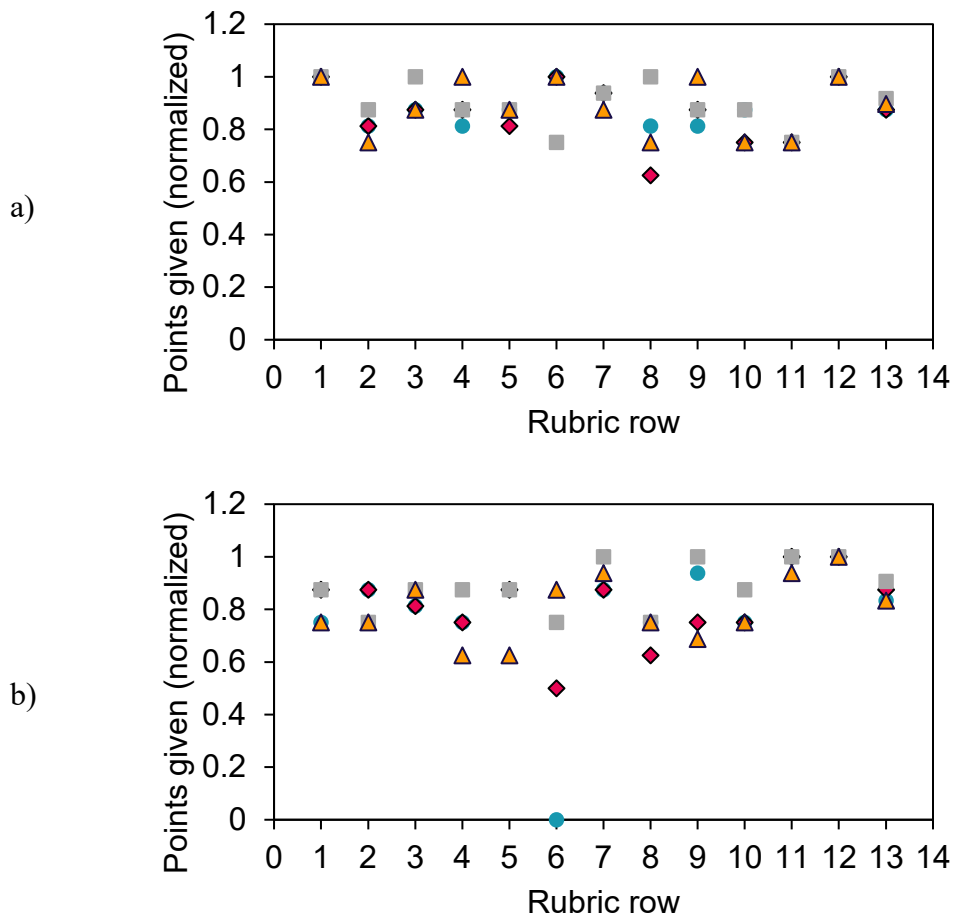


Figure 4: Scores from each of the four graders for student reports a) report 1 and b) report 3 from the fall rubric validation. The rubric row numbers on the x-axis correspond to the 13 rubric categories as listed in Table 1 and the symbol colors correspond to the grader colors in Figure 3.

For the IRR on the fall data, overall, the raters exhibited good consistency in their ratings of the four students: 9 of the 13 constructs had ICC’s that were excellent (Table 4), with Page Maximum showing near perfect consistency. The Overall Report construct was the combination of the three previously rated poor constructs (Tech Writing, Format, and Figure/Table Format) from the spring validation, and was rated as “excellent,” which confirms that there was less variability when the report is graded more holistically or potentially with fewer constructs.

Additionally, the ICCs of the total scores was excellent at 0.94. One construct, safety, had an ICC that was rated as good and 3 of the 13 constructs had poor consistency: Conclusions, Results, and Abstract. Interestingly, all of these were rated as excellent in the first rubric validation and had no changes made. The three constructs had variability in the ratings and, in going from one report to the next, the raters tended to go in opposite directions, with some raters going higher in their ratings while some going lower.

Table 4: Results of the fall validation IRR analysis as well as IRR analysis done by systematically excluding each grader. Boxes highlighted in green indicate an increased ICC after that grader was eliminated.

Construct	ICC value	Rating	w/o Teal grader	w/o Pink grader	w/o Gray grader	w/o Orange grader
Cover page	0.90	excellent	0.80	0.89	0.89	0.91
Abstract	0.00	poor	0.00	0.00	0.89	0.00
Intro/background/theory	0.90	excellent	0.84	0.86	0.86	0.92
Objectives	0.86	excellent	0.76	0.80	0.84	0.89
Method	0.81	excellent	0.68	0.80	0.68	0.97
Safety	0.72	good	0.49	0.38	0.71	0.79
Results	0.00	poor	0.00	0.00	0.18	0.00
Discussion	0.79	excellent	0.63	0.71	0.91	0.68
Conclusions/Rec	0.00	poor	0.00	0.00	0.23	0.00
References	0.80	excellent	0.82	0.67	0.77	0.71
Appendices	1.00	excellent	0.99	0.99	0.99	1.00
Page Maximum	1.00	excellent	1.00	1.00	1.00	1.00
Overall Report	0.84	excellent	0.72	0.72	0.85	0.89
Total Score	0.94	excellent	-	-	-	-

In order to determine if removing a single grader would improve the ICC of the constructs, the ICC was calculated by systematically eliminating a grader. The eliminated grader was then returned to the analysis and another grader eliminated until all four graders had been systematically removed. The last four columns of Table 4 list the results, with the ICCs that were improved upon eliminating that grader being highlighted in green. Removing the pink and teal graders tended to result in lower ICCs, which is not desirable. It should be noted that the pink and teal graders have taught a senior laboratory class over ten times each, assisted in developing the rubric with the Writing Center, and have a great deal of experience grading lab reports. Therefore, when they were removed from the ICC analyses, the overall consistency was lost and the ICC was diminished. Removing the gray and orange graders resulted in five and six increased constructs, respectively. The increased consistency found when the gray and orange graders were removed from the analysis indicates that these two graders may have had a debilitating effect on consistency. The gray grader had been an instructor on the senior lab course previously, while the orange grader was a new instructor. With the removal of the orange grader, the Methods construct went from good to excellent and after removing the gray grader, the Abstract construct went from poor to excellent. Although improved from an ICC of 0.00, the Results and Conclusions/recommendations constructs remained poor after removing the gray grader.

It is important to note that even with three of the constructs being rated as “poor,” the ICC of the total scores was 0.94, which is excellent. This indicates that even though there may be less consistency within the individual constructs, the overall report scores are consistent. Based on discussion with graders, there are likely several reasons for this. During the spring rubric validation discussion, graders noted that they tended to ensure the final report score was close to their overall thoughts on the report. That is, if the report they score receives a C grade, but overall, the instructor felt it was low B work, they revise their scores within the rubric. Additionally, some instructors have higher expectations for certain constructs, such as results or safety, and lower expectations for the abstract or introduction or vice versa. This was due to aspects like the instructor’s area of research expertise, their teaching experience, and their employment experience. One instructor with a statistics background due to their experience in industry placed strong importance on the student’s accurate use and interpretation of statistics. Another instructor with a theoretical research background was more concerned with the student’s ability to relate their analysis and interpretation to theoretical concepts. With this perspective, lower ICC values in individual constructs is not unexpected and indeed it is known that a rubric with fewer constructs is generally going to have better inter-rater reliability[12]. It seems that experienced graders have a good understanding of what constitutes quality technical communication and when grading in a holistic way are more likely to award consistent scores.

In order to test the hypothesis that experienced graders may be better calibrated in terms of the rubric grading, an additional five reports from different labs (a fuel cell and ionic diffusion experiment) were graded by the teal, pink, and gray grader. These five report scores were combined with the eight they had graded previously in the fall and spring validations and an IRR was completed (Table 5). Note that since the “Overall” construct was not in the spring rubric, an ICC was not determined for it on the complete dataset. It was found that five constructs, Objectives, Method, Appendices, Page Maximum, and Total Score, had excellent ratings, the cover page and references were good, and five constructs fell in the fair to poor category. It is worth noting that only the Results construct resulted in a poor rating, which is an improvement over previous validation.

Table 5: Results of IRR analysis on 13 reports graded by three experienced graders.

Construct	ICC value	Rating
Cover page	0.671	good
Abstract	0.417	fair
Intro/background/theory	0.46	fair
Objectives	0.837	excellent
Method	0.978	excellent
Results	0.362	poor
Discussion	0.451	fair
Conclusions/Rec	0.433	fair
References	0.688	good
Appendices	0.949	excellent
Page Maximum	1.00	excellent
Total Score	0.765	excellent

Looking more closely at the constructs that were rated fair to poor, there are clear areas of overlap between these constructs. For example, in the case of the Results construct and Discussion construct, many students combine these sections within the technical report. Additionally, some students will include the recommendations in the conclusions section and not the discussion (or vice versa). Similarly, the Intro/background/theory and the Objectives constructs have overlap as well. Students are encouraged, but not required, to have a separate Objectives section and sometimes include them in the introduction. Different instructors, as discussed previously, may value certain aspects over others or do “double counting” of error, leading to variation of scoring in these overlapping constructs. In order to test the hypothesis that overlapping constructs are impacting reliability, an IRR was done on the data in which the scores of the Discussion, Results, and Conclusions constructs were combined. This resulted in an ICC of 0.597, which has a rating of good for IRR (Figure 5). By combining the Intro/background/theory and Objectives constructs, the ICC was 0.715, resulting in a rating of good (Figure 5). If these changes were made, only one construct, Abstract, had a rating below good (Table 6).

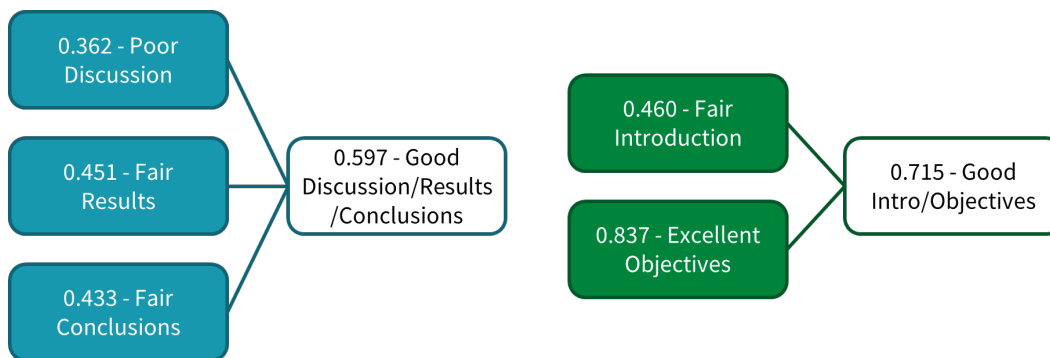


Figure 5: The overlapping constructs that were combined and the resulting ICCs.

Table 6: Results of the IRR analysis on 13 reports graded by three experienced instructors with the overlapping constructs combined.

Construct	ICC value	Rating
Cover page	0.671	good
Abstract	0.417	fair
Intro/background/theory/Objectives	0.715	good
Method	0.978	excellent
Results/Discussion/Conclusions	0.597	good
References	0.668	good
Appendices	0.949	excellent
Page Maximum	1.00	excellent
Total Score	0.765	excellent

In general, reducing the number of constructs in the rubric resulted in higher grader consistency. However, there are advantages to retaining the level of detail in the rubric with more constructs. A rubric such as used in this work, which explicitly describes in detail the multiple aspects of what quality work looks like for each construct, gives students a better understanding of what is

expected of them and allows them to clearly connect instructor comments to the construct on which they are being evaluated. Instructors also have a structure from which to formulate feedback and, in addition to discussions about scoring, this helps to keep standards consistent. Training or instructor orientations however are necessary to align expectations and make sure that feedback is not contradictory or that instructors do not have vastly different quality standards. A key take-away is also to ensure that the overall scores are balanced with the understanding that there may be some inconsistency within the individual constructs. Since the rubric has been implemented in this course, the instructors noticed a decrease in student evaluation comments critical of disparity in grading. Historically, these comments appeared even when instructor average scores were similar and consistent with class averages.

Conclusions

In conclusion, while a greater number of constructs may result in lower inter-rater reliability for individual constructs, they provide students and instructors with more detailed descriptions of expectations, making them a valuable learning tool for students and a way for instructors to be more deliberate in their grading process. This study revealed that consolidating constructs that may result in points being deducted across multiple categories for the same error into combined categories and a holistic category improved grader consistency. The same detailed descriptions of the expectations can still be included for the combined constructs in order to retain the rubric's value as a learning tool.

Additionally, providing instructors 'training' or 'orientation' on rubric interpretation, effective feedback, and rubric use before the course begins is recommended. This can ensure that all instructors are grading equivalently, and that overall scores have good reliability. This would consist of discussion on how to interpret each of the rubric constructs, discussion/instruction on how to give effective feedback and practice using the rubrics prior to the course. In laboratory courses with multiple instructors, it is also recommended to periodically check that score averages are consistent, and that feedback is effective and not contradictory. In addition, a sample report, i.e. a 'grading key' could be provided to instructors to compare to student work and graders could be asked to rank the constructs in order of importance in order to make grading biases visible. Curriculums with multiple courses with technical communication components could also benefit from instructors taking time to meet for alignment of standards.

Finally, we recommend preparing students to understand that technical communication is a complex process without definitive answers and that feedback may differ depending on who is grading. Students in laboratory courses with multiple instructors have the benefit of receiving feedback from multiple perspectives, allowing them to practice navigating shifting expectations in a relatively low stakes environment. Explicitly discussing this with students will help them understand that improving technical communication skills is a lifelong process and prepares them for navigating shifting expectations in their careers.

Acknowledgements

This work is supported by the National Science Foundation under Grant number 212077. Any results expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors also acknowledge the Norm Asbjornson College of Engineering's Bryan Innovative Instructional Grant Program and the MSU Writing Center's

Integrating Writing into STEM Grant Program for support. Special thanks to Writing Center Director Michelle Miley for meaningful discussions.

References

- [1] C. D. Grant and B. R. Dickson, "Personal Skills in Chemical Engineering Graduates: The Development of Skills Within Degree Programmes to Meet the Needs of Employers," *Education for Chemical Engineers*, vol. 1, no. 1, pp. 23-29, 2006.
- [2] M. C. Paretti, L. D. McNair, and J. A. Leydens, "Engineering Communication," in *Cambridge Handbook of Engineering Education Research*, A. Johri and B. M. Olds, Eds. Cambridge: Cambridge University Press, 2014, pp. 601-632.
- [3] P. Sageev and C. J. Romanowski, "A Message from Recent Engineering Graduates in the Workplace: Results of a Survey on Technical Communication Skills," *Journal of Engineering Education*, vol. 90, no. 4, pp. 685-693, 2001.
- [4] C. A. Hubka *et al.*, "A Writing in the Disciplines Approach to Technical Report Writing in Chemical Engineering Laboratory Courses," presented at the ASEE Annual Conference, Tampa, Florida, 2019. Available: <https://peer.asee.org/32019>
- [5] D. Miller and J. Williams, "Incorporating Peer Review Into The Che Laboratory," presented at the ASEE Annual Conference, Salt Lake City, Utah, 2004. Available: <https://peer.asee.org/13191>
- [6] K. Wright and P. E. Slaboch, "Board 100: Enhancement of a Thermo-Fluid Laboratory Course: Focus on Technical Writing," presented at the ASEE Annual Conference, Tampa, Florida, 2019. Available: <https://peer.asee.org/32165>
- [7] R. A. Fenner and P. O'Neill, "Board 44: Work in Progress: Integrating Writing into Engineering Labs: Developing Curriculum and Creating a Writing Fellows Program," presented at the ASEE Annual Conference, Tampa, Florida, 2019. Available: <https://peer.asee.org/32351>
- [8] J. S. Underwood and A. P. Tregidgo, "Improving Student Writing Through Effective Feedback: Best Practices and Recommendations," *Journal of Teaching Writing*, vol. 22, pp. 73-98, 2010.
- [9] S. G. Wettstein and J. R. Brown, "Using Existing University Resources: Integration of the University Writing Center into a Senior-level Laboratory Series for Improved Learning Outcomes," presented at the ASEE Annual Conference, Virtual Conference, 2021. Available: <https://peer.asee.org/37992>
- [10] R. K. Boettger, "Rubric Use in Technical Communication: Exploring the Process of Creating Valid and Reliable Assessment Tools," *IEEE Transactions on Professional Communication*, vol. 53, no. 1, pp. 4-17, 2010.
- [11] S. Sheffield, R. Fowler, L. Alford, K., and K. Snyder, "Implementing a Single Holistic Rubric to Address Both Communication and Technical Criteria in a First Year Design-Build-Test-Communicate Class," presented at the ASEE Annual Conference, Columbus, Ohio, 2017. Available: <https://peer.asee.org/28479>
- [12] A. Jonsson and G. Svingby, "The use of scoring rubrics: Reliability, validity and educational consequences," *Educational Research Review*, vol. 2, no. 2, pp. 130-144, 2007.